



**SOLID STATE
ELECTRONICS
LABORATORY**

JSEP ANNUAL REPORT

1 March, 1994 through 28 February, 1995

**James S. Harris, Jr.
JSEP Principal Investigator
and Program Director**

(415) 723-9775

**This work was supported by the
Joint Services Electronics Program
(U.S. Army, U.S. Navy and U.S. Air Force)
Contract DAAH04-94-G-0058,
and was monitored by the
U.S. Army Research Office**

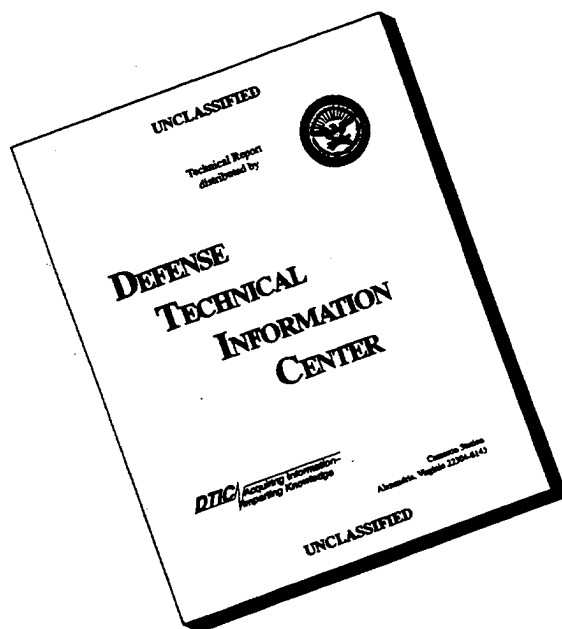
DTIC QUALITY INSPECTED 4

**Reproduction in whole or in part is permitted
for any purpose of the United States Government**

**This document has been approved for public
release and sale; its distribution is unlimited**

19960212 016

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

JSEP ANNUAL REPORT

1 March, 1994 through 28 February, 1995

**James S. Harris, Jr.
JSEP Principal Investigator
and Program Director**

(415) 723-9775

**This work was supported by the
Joint Services Electronics Program
(U.S. Army, U.S. Navy and U.S. Air Force)
Contract DAAH04-94-G-0058,
and was monitored by the
U.S. Army Research Office**

**Reproduction in whole or in part is permitted
for any purpose of the United States Government**

**This document has been approved for public
release and sale; its distribution is unlimited**

Abstract

This is the annual report of the research conducted at the Stanford Electronics Laboratories under the sponsorship of the Joint Services Electronics Program from March 1, 1994 through February 28, 1995. This report summarizes the area of research, identifies the most significant results and lists the dissertations and publications sponsored by the contract DAAH04-94-G-0058.

Table of Contents

| | |
|---|----|
| Introduction | 2 |
| Unit 1: Investigation of Transport in Quantum Dots | 4 |
| Unit 2: Physics and Applications of Ultra-small Structures | 13 |
| Unit 3: Investigation of a Metal Source and Drain Field Emission Transistor | 23 |
| Unit 4: The Electronic Structure and Interfacial Properties of High Temperature Superconductors | 33 |
| Unit 5: On-chip Thin Film Solid State Micro-battery | 39 |
| Unit 6: CVD Epitaxial Germanium <i>n</i> -channel FETs Formed on Si Substrates using Strain-relief Layers | 43 |
| Unit 7: Portable Video on Demand in Wireless Communication | 51 |
| Unit 8: Wideband Data Transmission Techniques | 61 |
| Unit 9: Smart Antennas for Mobile Radio Networks | 69 |
| Unit 10: Efficient Data Compression | 79 |

This work was supported by the Joint Services Electronics Program, contract DAAH04-94-G-0058. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies either expressed or implied of the U.S. Government.

JSEP ANNUAL REPORT

January 11, 1994 - January 10, 1995

Introduction

The JSEP contract supports a program of unclassified basic research in electronics conducted by faculty members of the Electrical Engineering Department of Stanford University as a component of the research program of the Stanford Electronics Laboratories. The Stanford Electronics Lab JSEP Director and Principal Investigator is Professor James Harris. He is responsible for the selection of the best individual proposals, coordination between Stanford and the JSEP TCC and coordination between the selected areas of the JSEP Program. In planning the JSEP Program at Stanford, a general objective is to develop new projects with 3-4 years of JSEP sponsorship, leading to a transition to more conventional DoD or other agency program funding. Since this type of funding often requires 12-18 months in the proposal, evaluation and budgeting stages, the flexibility in JSEP funding allows us to seize new opportunities and initiate programs which might otherwise be delayed for a significant period. Following this course, three projects from the earlier program now have additional DoD funding. Examples of funding initiation on the current program are: the initiation of a project to create precisely controlled Si nanostructures and examine their light emission properties and the initiation of project on sub-micron patterning of magnetic thin films for high density storage. Both of these projects have achieved significant results in a short time that would have been impossible to achieve without the flexibility afforded by the Director's ability to change project support with JSEP funding.

The technical knowledge developed under the JSEP contract is widely disseminated through sponsor reviews, presentations of papers at technical meetings, publications in the open literature, discussions with visitors to the laboratories, and publication of laboratory technical reports (Ph.D. dissertations). Major successes of four projects are highlighted below.

Highlights

Sufficiently small magnetic particles contain no domain walls and are uniformly magnetized. A magnetic recording medium consisting of an array of single-domain islands would be ideal for storage of a noise free single bit of information. Patterned polycrystalline magnetic thin films using direct-write electron beam lithography and a multi-step masking and ion milling process were used to define large arrays of various sized cobalt islands. A transition from a multi-domain to a

single domain state was demonstrated at an island diameter of roughly $0.2\mu\text{m}$. This offers a significant advance in magnetic storage.

One of the keys to application of quantum dots will be to use them in large arrays and utilize their unique switching and hysteretic properties in some novel way. We have fabricated and investigated the properties of the largest array (200×200) of quantum dots. This array was possible by a combination of e-beam lithography, etching and a novel single Schottky gate. Comparing the switching behavior between a single quantum dot and the above array, we observe no hysteresis in the single dot compared to the many steps in the I-V characteristic of the array, each step with a hysteretic loop. We believe this unexpected and previously unexplained hysteretic behavior is due to the Schottky gate formed at the sides of the quantum dots. We are currently seeking to confirm this hypothesis by investigating the importance of the gate-to-2DEG tunneling in the observation of hysteretic behavior.

Power consumption in the digital signal processing element is the key design element for improved portable video-on-demand. Memory access is by far the most power consuming operation, thus the main strategy in our approach was minimizing memory access. A pyramid vector quantization decoder chip was combined with a subband decoder for real-time video decompression. This custom IC approach provided a 100 X reduction in power compared to a conventional C-Cube JPEG decoder (not including the additional memory power for the JPEG decoding). We demonstrated that such a hardware-driven algorithm design strategy can deliver high-quality video at an extremely low power level.

One of the central problems for mobile radio networks is the capacity (number of users per cell) and quality (outage probability) of such networks. Limiting factors in the capacity and quality of current wireless systems are the availability of RF spectrum, which requires more efficient spectral usage. Using antenna array processing algorithms, we have demonstrated smart transmitters that use information on the mobile unit locations to emit directional radiation towards the intended mobile unit. Similar approaches for the receiver enhance the received signal. These techniques greatly reduce the mutual interference and allow multiple co-channel user within a single cell, which increases capacity several fold.

Unit: 1

TITLE: Investigation of Transport in Quantum Dots

Principal Investigator: J. S. Harris, Jr.

Graduate Students: C. I. Duruöz and D. Stewart

1. Scientific Objectives

With recent developments in e-beam lithography techniques, it has become possible to fabricate devices with minimum feature sizes on the order of 100nm. Two dimensional electron gas(2-DEG) structure implemented with GaAs technology is an exceptionally convenient system for studying these nanoscale devices. A particularly interesting nanoscale device is called "Quantum Dot", a submicron region of a semiconductor which confines some number of electrons. In such small quantum dots, one may see significant electron-electron interactions and charging effects which are central to much of the recent experimental [Kouwenhoven, Johnson, McEuen] and theoretical [Meir, Beenakker] work on semiconductor quantum dots. A very interesting and experimentally accessible system is a two dimensional array with adjustable coupling between the array elements. As discussed in recent theoretical studies [Middleton, Stafford], and as we will demonstrate experimentally in this report, it is possible in this system to investigate possible collective effects in transport and their relation to the strength of dot-to-dot interactions. In particular, it has been predicted that arrays of quantum dots show a threshold for conduction due to the effects of disorder and Coulomb blockade [Middleton].

2. Progress and Experimental Results

2.1 Introduction

We performed experiments on two dimensional quantum dot arrays where a single gate is used to form and control the barriers between the individual elements, as well as to change the density of the two dimensional electron gas (2DEG) (Fig. 1 (A) (B)). The current-voltage (I-V) characteristics of the arrays have two main features: a threshold for conduction, and multiple switching events accompanied by hysteresis. Multiple switching events result in a hierarchy of hysteresis loops in the arrays. By changing the gate voltage, V_g , it is possible to move between

the hysteretic and non-hysteretic regimes. In a control dot fabricated on the same chip, we also observe a single hysteresis loop accompanied by a single switching event, in contrast to the array. This is different than the behavior of most top-gated quantum dots studied so far, an exception being the hysteresis observed by Wu *et al.* [Wu] in double barrier lateral structures. In this case, hysteresis was attributed to electron heating once the current begins to flow through the device [Goodnick]. Because the hysteresis and switching in our arrays occurs for currents as low as 1.5 nA at voltages of 3.0 mV, we believe that electron heating is not the cause of hysteresis in our case. Instead, we propose that the hysteresis is associated with charge exchange in the form of a small leakage current to the gate in these structures. Since the I-V curves were repeatable within a cool down and stable during very slow sweeps, we believe these effects are not related to changes in occupation of impurity states.

2.2 Device Fabrication and Measurement Setup

Arrays of 200 by 200 dots were fabricated using a standard modulation doped GaAs/Al_{0.34}Ga_{0.66}As 2DEG structure with an electron mobility of $\sim 200,000$ cm²/Vs and a sheet density of 3.5×10^{11} cm⁻² at 4.2 K. The 2DEG layer is 770 Å below the surface, consisting of a 300 Å AlGaAs spacer above the 2DEG, a 170 Å Si doping layer ($N_D = 3.8 \times 10^{18}$ cm⁻³) and a 300 Å undoped GaAs cap layer. To form the dots, "plus sign" patterns with a spatial period of 0.8 μ m (Fig. 1 (B)) were formed by electron-beam lithography and subsequent wet-etching roughly 800 Å

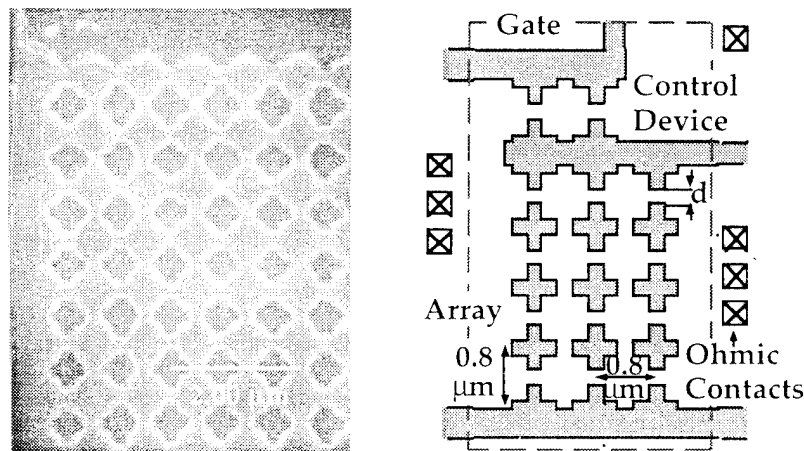


Figure 1. (A) Electron micrograph of part of the 200 x 200 array. Electron gas is removed beneath darker regions; channels between dots can be depleted with negative gate voltage in the -40 to -150 mV range depending on the device. (B) Schematic diagram showing the layout of the array, control device and ohmic contacts. The lithographic distance "d" is 300, 250 and 200 nm for device 1, 2 and 3, respectively.

deep, through the 2DEG layer. Three devices were measured (referred to as "1", "2" and "3") with spacings of the lithographic distance between closest edges of the plus signs of 300 nm, 250 nm, 200 nm, respectively. The devices were measured in a dilution refrigerator at mixing chamber temperatures between 20 and 700 mK. A slowly swept dc voltage bias plus a 10 μ V, 11.4 Hz ac voltage bias were applied across the array.

In most single-dot experiments, tunnel barriers coupling the dot to the bulk 2DEG are controlled by independent gates with an additional gate used to adjust the electron density [Wu]. This strategy is difficult to implement in a two-dimensional array. Instead we use a single Cr/Au gate deposited over the entire array as well as over an isolated control dot which has the same dimensions as each individual array element (Fig. 1). A small negative gate voltage ($|V_g| < \sim 300$ mV, typically) constricts the barriers by increasing the lateral depletion and raising the potential in the barrier. This greatly increases barrier resistances between the dots without producing large density changes within the dots (roughly 630 mV is necessary to fully deplete the 2DEG).

2.3 Experimental Results

Figure 2 shows typical I-V curves as a function of gate voltage, V_g . These curves illustrate a type of multiple hysteresis loops observed as a function of the inter-dot coupling, which is adjusted by the gate. For $V_g = -98$ mV, the I-V curve has a single loop near 4 mV bias voltage. As the gate voltage becomes more negative, the width of this loop increases and a new hysteresis loop appears for $V_g < -106$ mV. These two loops merge at a gate voltage between -114 mV and -118 mV. Figure 2 also illustrates the discontinuous jumps in the current (within the resolution of a single data point), which we will refer as "switching events". In the curves for $V_g < -118$ mV, multiple switching events occurring in a single loop and can be very clearly noticed. In all hysteresis loops observed, the switching-on voltage for increasing V_{arr} is larger than the switching-off voltage when V_{arr} is decreased, that is; all hysteresis loops are counter-clockwise in I versus V. We also find sub-loops on both the upper and lower parts of the curve if the sweep direction is reversed following a current jump. Hysteresis in these subloops is also always counter-clockwise in I versus V. Device 3, which has narrower channels between the dots compared to device 2, has a larger number of hysteresis loops in the I-V curve.

Figure 3 shows that switching voltages decrease for increasing temperature, and the width of each hysteresis loop also decreases as the temperature increases. It is possible to see in Fig. 3 the dissociation of big loops into smaller ones and their disappearance at different temperatures, all

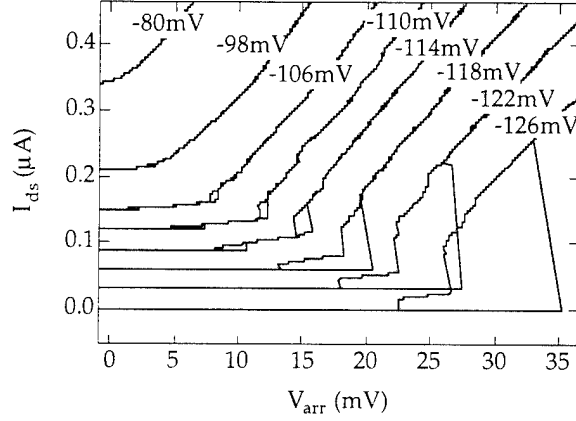


Figure 2. Typical I-V curves for several gate voltages (device 2) measured at base temperature ($T \sim 20\text{mK}$). The curves are offset in proportion to gate voltage (V_g) for clarity. V_{arr} is the voltage measured directly across the array in a four-lead configuration.

lower than $\sim 700\text{ mK}$. There is also an apparent tradeoff between gate voltage and temperature: at 700 mK , the hysteresis can be recovered if the gate voltage is made 20-30 mV more negative. However, the ratio of the loop width to the switching voltage is always smaller than that at 20 mK . This suggests that switching and hysteresis will inevitably disappear at sufficiently high temperatures regardless of gate voltage. Indeed, at 4.2 K , we observe no hysteresis at all in any of the samples.

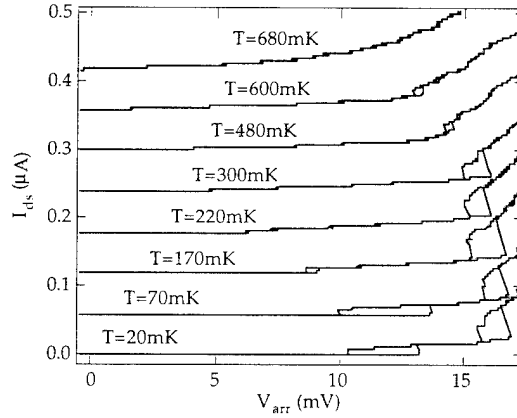


Figure 3. I-V curves (offset for clarity) for device 2 at $V_g = -115\text{ mV}$ at various temperatures.

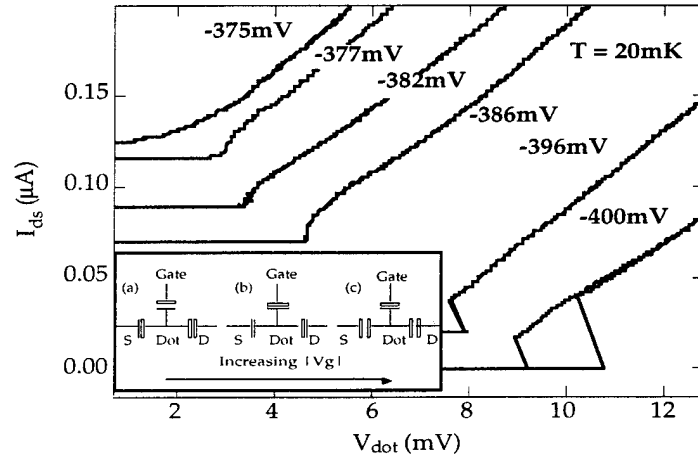


Figure 4. I-V curves for the single dot (control device 2), at various gate voltages, V_g . The curves are offset in proportion to gate voltage for clarity. The inset is the schematic illustration of expected change in dot-dot and dot-gate transparencies as a function of gate voltage. At $V_g=0$, the resistive coupling between dots is large. As the gate voltage becomes more negative dot-dot transparency decreases (a), and after a crossover (b), the gate-dot transparency can become greater than the dot-dot transparency (c). In this last situation, there is non-negligible resistive coupling between the gate and dots.

We have also investigated the I-V properties of the single control device located adjacent to the array on each sample. The same experimental measurement configuration as for the arrays was used to obtain the results shown in Fig. 4. In the case of a single dot, we observe no hysteresis near pinch-off ($V_g \sim -375$ mV), however, beyond a gate voltage ~ 20 mV more negative than this pinch-off value, a single hysteresis loop appears, accompanied by upward and downward switching events. Here, we define pinch-off as the regime where the device (array or single dot) has negligible conductance near zero bias, i.e., the I-V curve has zero slope at the origin. The I-V curve for the single dot has a very weak temperature dependence compared to the array, and the width and location of the hysteresis are unchanged up to 700 mK. As in the array, no hysteresis is seen at 4.2 K for any gate voltage. Unlike the arrays, no sub-loops or multiple switching events are observed in the single dot.

2.4 Discussion of the Results

To address the cause of the hysteretic behavior in our experiments, we consider a model of a semiconductor quantum dot that includes a weak resistive coupling from the 2DEG to the gate as shown in Fig. 4 inset. We believe this coupling is particularly important in our samples because the dots are formed by wet etching with a gate that fills in the etched regions. This allows high

resistance (typically $\sim 10^8 \Omega$) barriers to form along the sidewalls of the dots. At $V_g = -100$ mV, a gate current of ~ 75 pA is measured for the deep etched devices (1, 2, and 3) reported here. In other shallow-etched samples, which in general did not show hysteresis, a typical gate current of ~ 1 pA was measured at the same gate voltage. The mechanism by which a conducting path from the gate to the dot can induce hysteresis has been analyzed recently in a model of a Single Electron Transistor (SET) coupled to a controlling potential (V_g here) through a series resistor R_0 and a capacitor, C_g [Korotkov]. This model, known as the RC-SET, gives hysteresis in the limit of large coupling resistance to the controlling potential. In making this comparison, our gate-to-dot resistance $R_0(V_g)$, which increases with more negative V_g in our devices, corresponds to the coupling resistance R_0 in Ref. [10]. Experimental parameters for both the arrays and the single dots ($R_0(V_g) > \sim 10^8 \Omega$, tunnel barriers $R_1(V_g), R_2(V_g) > \sim h/e^2$) are in the range where the RC-SET model predicts hysteresis, $R_0(V_g) \gg (R_1 + R_2) C_S/C_g \gg h/e^2$ (C_S is the total capacitance of the dot). This model suggests that a dot in an array that in isolation would not be hysteretic, can be pushed into the hysteretic regime by the impedance of its neighbors, which effectively raises R_1 and R_2 for that dot.

3. Future Work

To study the cause of the hysteresis, we have been fabricating single dots with three leads (Fig. 4). The devices are made using a similar 2-DEG structure as in the arrays and e-beam lithography technique combined with subsequent lift-off. The goal is to use the third lead as an additional channel which allows some charge leakage into the dot, in a very controlled way and therefore simulate the effect of the gate leakage in the arrays. All the barrier transparencies can be independently adjusted by means of independent gates (four in Fig. 4). The devices will be measured in a dilution refrigerator with the same experimental measurement configuration as that used for the measurement of the arrays.

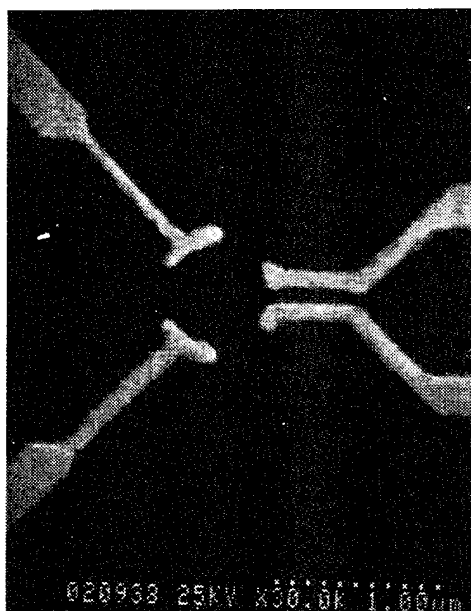


Figure 5. The SEM Picture of a single dot with three leads

We are also planning to fabricate one dimensional and small two dimensional arrays by using the e-beam lithography and wet etching techniques as in the 200x200 arrays.

4. References

- [Beenakker] C. W. J. Beenakker (1991). *Phys. Rev. B* **44**, 1646-1656.
- [Goodnick] S. M. Goodnick, J. C. Wu, M. N. Wybourne and D. D. Smith (1993) *Phys. Rev. B* **48**, 9150-9153.
- [Johnson] A. T. Johnson, L. P. Kouwenhoven, W. de Jong, N. C. van der Vaart, C. J. P. M. Harmans and C. T. Foxon (1992) *Phys. Rev. Lett.* **69**, 1592-1595.
- [Korotkov] A. N. Korotkov (1994) *Phys. Rev. B* **49**, 16518-16522.
- [Kouwenhoven] L. P. Kouwenhoven, A. T. Johnson, N. C. van der Vaart, A. van der Enden, C. J. P. M. Harmans and C. T. Foxon (1991) *Z. Phys. B - Condensed Matter* **85**, 367-373

- [McEuen] P. L. McEuen, E. B. Foxman, U. Meirav, M. A. Kastner, Y. Meir and N. S. Wingreen (1991) *Phys. Rev. Lett.* **66**, 1926-1929.
- [Meir] Y. Meir, N. S. Wingreen and P. A. Lee (1991) *Phys. Rev. Lett.* **66**, 3048-3051.
- [Middleton] A. A. Middleton and N. S. Wingreen (1993) *Phys. Rev. Lett.* **71**, 3198-3201.
- [Stafford] C. A. Stafford and S. Das Sarma (1994) *Phys. Rev. Lett.* **72**, 3590-3593.
- [Wu] J. C. Wu, M. N. Wybourne, C. Berven, S. M. Goodnick and D. D. Smith (1992) *Appl. Phys. Lett.* **61**, 2425-2427.

JSEP Supported Publications

1. C. I. Duruöz, R. M. Clarke, C. M. Marcus and J. S. Harris, Jr., "Conductance Threshold, Switching and Hysteresis in Quantum Dot Arrays," *Phys. Rev. Lett* (in press)
2. C. M. Marcus, R. M. Clarke, I. H. Chan, C. I. Duruöz and J. S. Harris, Jr., "Phase-breaking rates from conductance fluctuations in a quantum dot," (1994) *Semiconductor Science and Technology* **9**, 1897-1901.
3. C. I. Duruöz, R. M. Clarke, C. M. Marcus and J. S. Harris, Jr., "Switching and Hysteresis in Quantum Dot Arrays." (to be published in the *Journal of Nanotechnology*, ed. by M. A. Reed)
4. R. M. Clarke, I. H. Chan, C. M. Marcus, C. I. Duruöz, J. S. Harris, Jr., K. Campman and A. C. Gossard, "Temperature Dependence of Phase Breaking in Ballistic Quantum Dots," (submitted to *Phys. Rev. Lett.*).

Unit: 2

TITLE: Physics and Applications of Ultra-small Structures

SENIOR INVESTIGATOR: R. F. W. Pease

RESEARCH STUDENT: R. M. H. New and H. Liu

Summary of Research

There are two main areas of research within this study:

- 1) Patterned thin film media for high density magnetic recording
- 2) Fabrication and light emission in Si nanocrystals

1. Patterned thin film media for high density magnetic recording

In conventional hard-disk magnetic recording systems, the signal to noise ratio is often limited by "transition" noise which occurs due to the irregular zig-zag domain walls between adjacent recorded bits [Tong]. In order to address this problem, we are studying recording media composed of large arrays of submicron lithographically defined single-domain magnetic islands.

It is known both from theoretical arguments and from experiments that sufficiently small magnetic particles are uniformly magnetized and contain no domain walls. If a single-domain particle of this type has a single uniaxial easy axis of magnetization then it will have only two possible magnetization states and will be ideal for storage of a single bit of information. A magnetic recording medium consisting of an array of equally spaced and uniformly shaped single-domain islands with predictably oriented easy axes could serve as a virtually noise-free alternative to the unpatterned magnetic thin films used in conventional hard disk systems. The ultimate theoretical storage density for such a system would be limited only by the spontaneous thermal switching of bits, a problem that would occur only for particles one hundred angstroms in diameter or less.

We have developed a procedure for patterning polycrystalline magnetic thin films using direct-write electron beam lithography and a multi-step masking and milling process [New (a)]. We have used this procedure to define large arrays of 0.15mm by 0.2mm cobalt islands. We have studied the physical properties of these islands using atomic force, scanning electron and transmission electron microscopy. The magnetic properties have been examined with both magnetic force microscopy and bulk hysteresis loop measurement techniques [New (b)].

For our initial experiments we have patterned magnetic islands out of a 200-Å-thick polycrystalline cobalt film. Our results indicate that the transition from the multidomain to single domain state occurs at an island diameter of roughly 0.2μm. Figure 1 shows an atomic force microscope image of an array of 0.2μm by 0.4μm islands. The magnetic force microscopy images of these islands show domain structure, so these islands are not single domain. However, smaller islands, roughly 0.15μm by 0.2μm in size, are almost all single domain. Figure 2 shows the characteristic dipole-like fields emerging from an array of these smaller uniformly magnetized particles.

Transmission electron microscopy images of the patterned polycrystalline islands indicate that there are roughly 200 cobalt grains per island, each of which has an easy axis of magnetization randomly oriented in the plane of the film. For islands with only a few hundred grains or less, the magnetocrystalline anisotropies of the individual grains may not completely average out and the net magnetocrystalline anisotropy may be larger than the shape anisotropy for some island geometries. Our calculations indicate that for the island geometries we are using, there is a significant probability that the net easy axis may be misaligned with the long axis of the island [New (c)], and our initial experiments confirm this. Such unpredictably oriented easy axes would cause problems in a single-bit-per-island recording scheme.

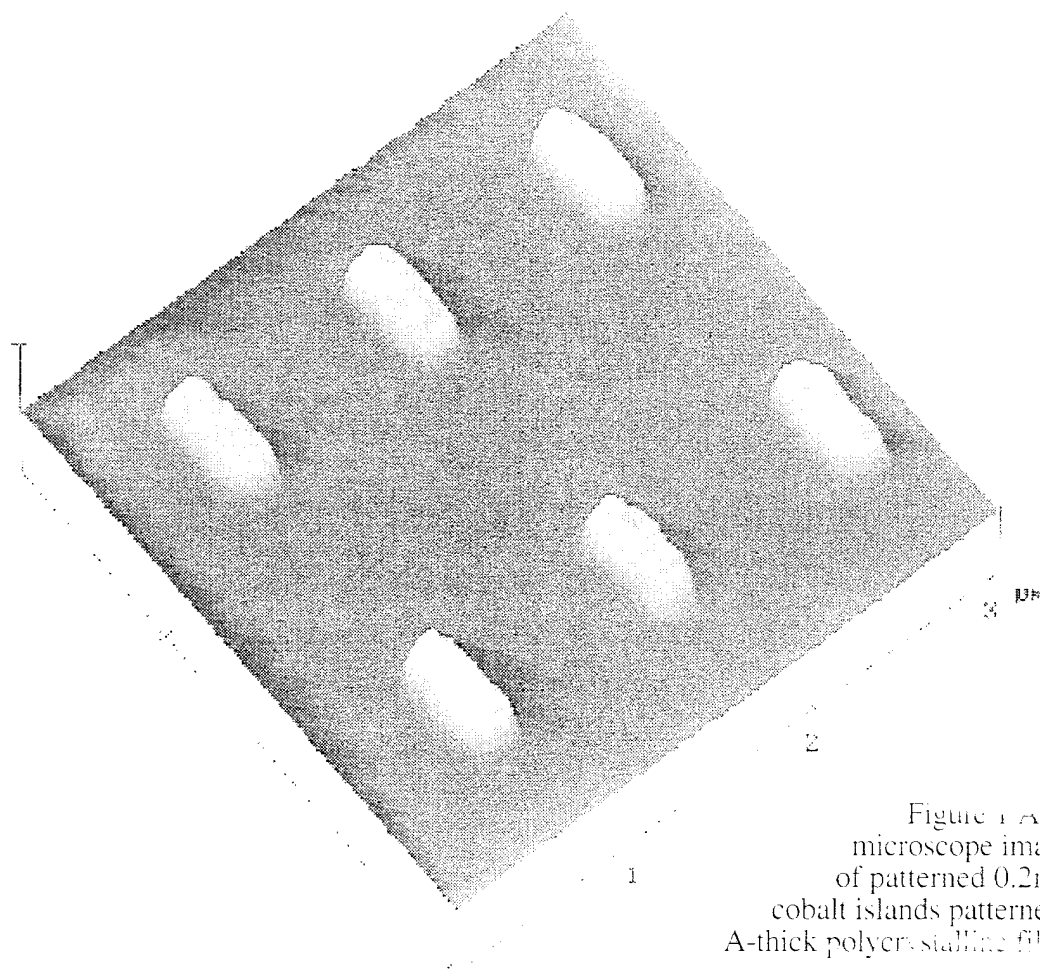


Figure 1 An atomic force microscope image of an array of patterned 0.2μm by 0.4μm cobalt islands patterned from a 200 Å-thick polycrystalline film

One problem with polycrystalline magnetic recording films, either patterned or unpatterned, is that the fundamental unit of magnetization (typically a single grain or grain cluster of 100 to 500 Å in diameter) is not much smaller than the size of a single recorded bit. For a state of the art 1Gbit/in² recording system, there may be only a hundred grain clusters or less per bit. Because the medium is so coarsely discretized, conventional magnetic recording systems suffer from increasing signal to noise problems as recording densities are increased. Recording medium noise is already the most important component of noise in recording systems which use magnetoresistive readback heads.

One solution to this problem is to switch to a recording medium which is homogeneous over the size range of a single recorded bit. Sputtered single crystal films would provide a more controllable and predictable magnetic behaviour when examined at this size range, and patterned islands of single crystal material would not suffer from the problem of randomly oriented easy axes discussed above. One of the major advantages of the patterning technique that we have developed is that, unlike a lift-off process, it can be used to pattern single crystal thin films. In preparation

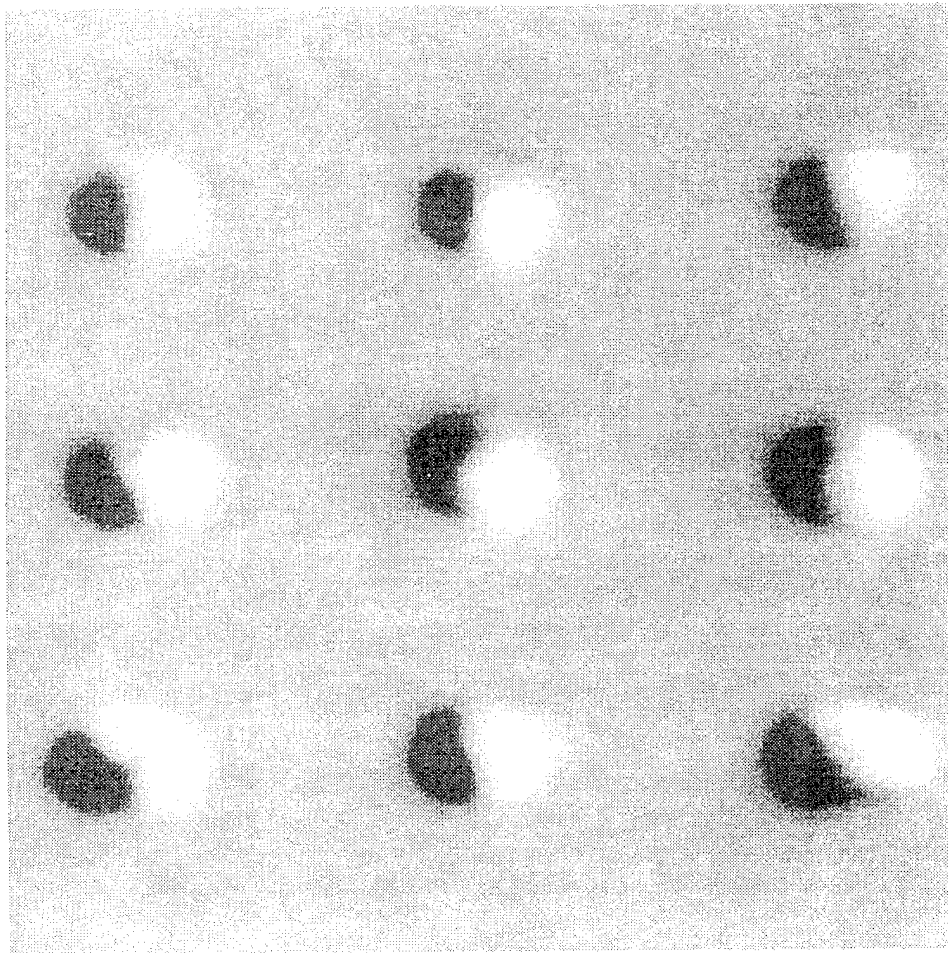


Figure 2. Magnetic force microscopy image of an array of 0.15mm by 0.2mm islands. These islands are uniformly or almost uniformly magnetized.

for future experiments we have sputter-deposited single crystal iron films on sapphire substrates and measured their magnetic and structural properties. These films show good epitaxial quality and have a predictably oriented uniaxial anisotropy as required.

In addition to the possible technological applications of patterned uniformly magnetized single crystal islands, such structures would be very interesting for theoretical reasons. There exists a body of micromagnetic calculations devoted to the most persistent question asked of fine ferromagnetic particles: how small must a ferromagnetic particle be in order for its lowest energy state to be single domain? The theoretical attempts to answer this question treat only uniformly shaped single crystal magnetic particles [Brown]. The possibility of making large arrays of uniformly sized and shaped single-domain such islands offers the possibility of experimental verification of these calculations.

2. Fabrication and Light Emission in Si Nanocrystals

The initial euphoria over the photoluminescence phenomenon of porous silicon [Canham] has gradually subsided. However, the interest in understanding the intrinsic properties of silicon nano-structures remains. Basic theoretical studies indicate that silicon nano-structures should

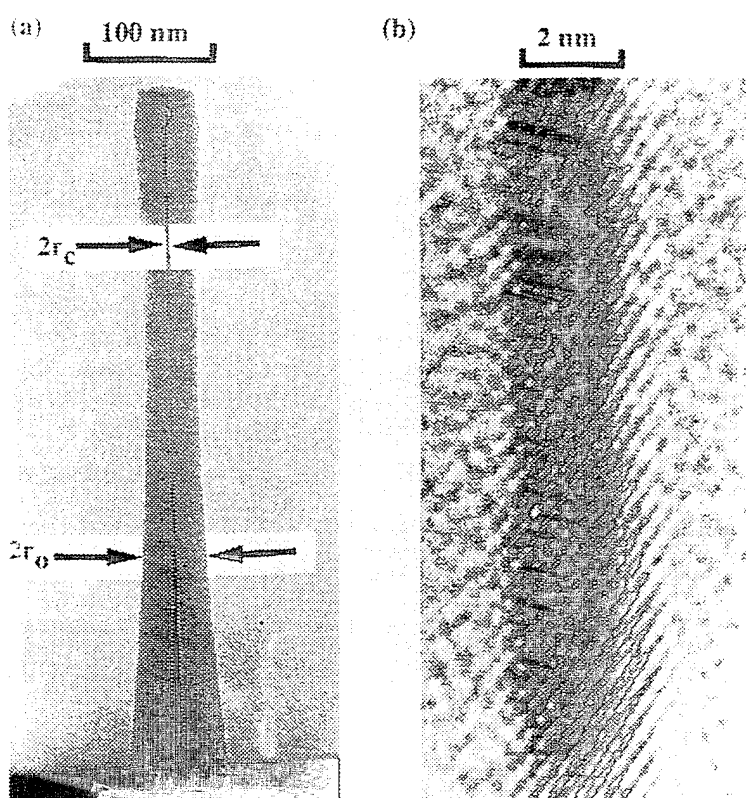


Figure 3. Side-view TEM micographs of an oxidized silicon nano-pillar reaching self-limiting regime. The silicon pillar was oxidized in dry oxygen at 875°C for 10 hours. The general shape of the entire oxidized column is shown in (a). A high resolution lattice image of the inner crystalline silicon core (2nm in diameter) is shown in (b). The lattice fringes are due to the {111} crystal planes with a separation of 3.14Å between consecutive planes.

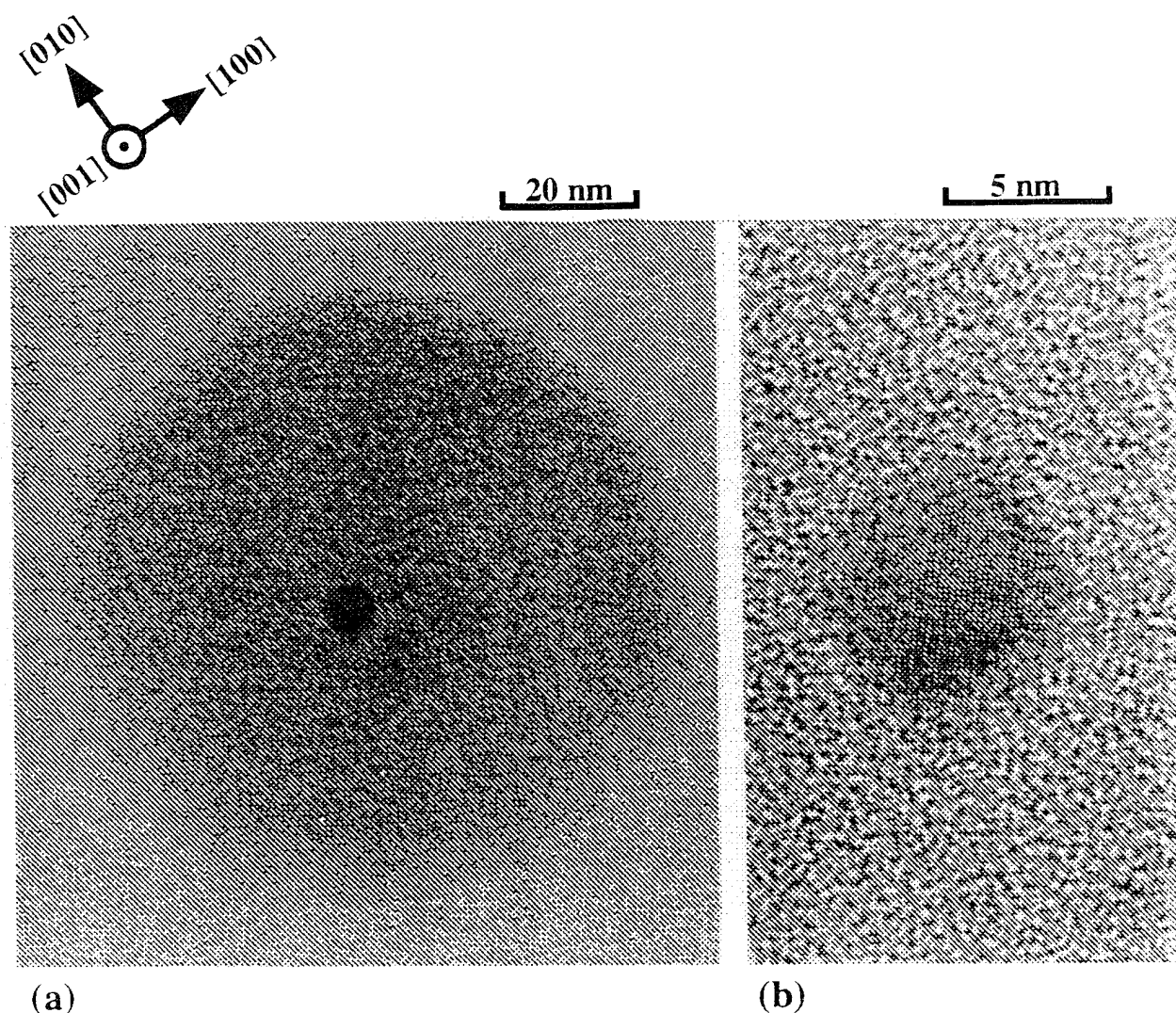


Figure 4. $\langle 100 \rangle$ top view cross-sectional TEM micrograph of Oxidized silicon Nano-pillar. The thin disk of the oxidized nano-pillar (1050°C dry oxidation for 1.5 hours) was prepared by a combination of polyimide planarization, mechanical polishing and ion milling. The pillar was patterned on a $\{100\}$ wafer (the crystal orientation marker is shown on the upper left hand corner). The dark crystalline silicon core (core width is 6 nm) surrounded by round amorphous oxide skin (outer diameter is 90 nm) due to diffraction contrast is clearly visible in (a). Expanded view of the silicon core in (b) shows the crystalline fringes due to the $\{220\}$ planes, which has a plane spacing of 1.92\AA . It also shows that the core facets in the $\langle 100 \rangle$ directions.

exhibit interesting and dramatically different electrical, optical, and thermal properties when compared to bulk silicon. Some of these properties include the enlargement of bandgap, room temperature carrier freeze-out, electron localization effects, pseudo-direct optical transitions, large energy shifts and fine structures in photon absorption spectra, and significant decrease in heat capacity and thermal conductivity. In our theoretical studies, we also found that these phenomena are not expected to become significant until the nano-structure dimensions are well below 10 nm. This places a great burden on the fabrication technology.

In the past few years, a process based on a combination of high resolution electron beam lithography, anisotropic reactive ion etching, and thermal oxidation was developed to fabricate well-controlled sub-5 nm silicon nano-wires [Liu]. A transmission electron microscopy (TEM) technique was also developed to examine the oxidized nano-wires with atomic resolution (see Fig. 3 for a 2 nm wide nano-wire). Recently, this TEM technique was adopted to study the unique self-limiting oxidation phenomenon in a systematic fashion [Liu]. It was also found through the top-view cross-sectional TEM technique that the nano-wires facet in low index plane directions when reach the self-limiting regime (see Fig. 4 for facets in $\langle 100 \rangle$ directions). In this regime, one obtains square nano-wires rather than circular ones. These represent the first detailed oxidation study of sub-10 nm silicon nano-structures.

A consistent oxidation model based on the theory of viscous flow was also developed to explain the extremely slow oxidation rate of silicon nano-wires [Liu] and nano-spheres [Okada]. Based on the parameters obtained from this model, a set of process curves was constructed to predict the oxidation time required for achieving sub-5 nm silicon nano-wires (Fig. 5 and Fig. 6). This can serve as a guide for future experimental study.

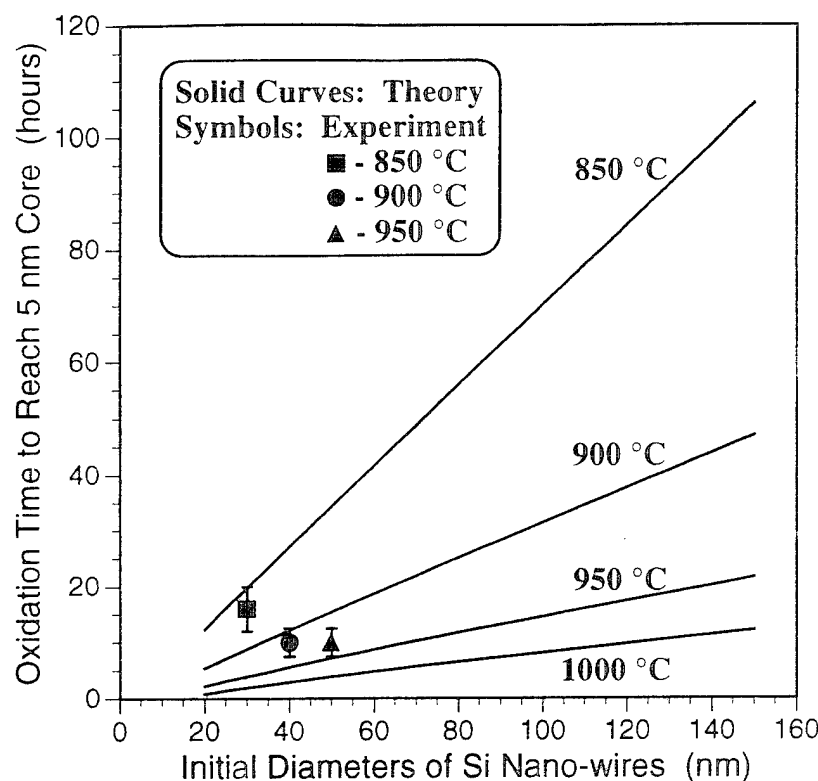


Figure 5. Dry oxidation time to reach 5 nm core vs. initial silicon column diameters. Using the set of viscous flow parameters obtained to fit the experimental data in this nano-wire oxidation study, the oxidation time required to yield 5 nm wide silicon cores was computed for a series of oxidation temperatures (850°C, 900°C and 1000°C). Since the viscous flow model is still no adequate to describe the even slower oxidation rate in the self-limiting regime, these calculations can be treated as the lower bound for the actual time to achieve the given core size. For comparison, three experimental data points were plotted in the same graph. Fits are within experimental accuracy.

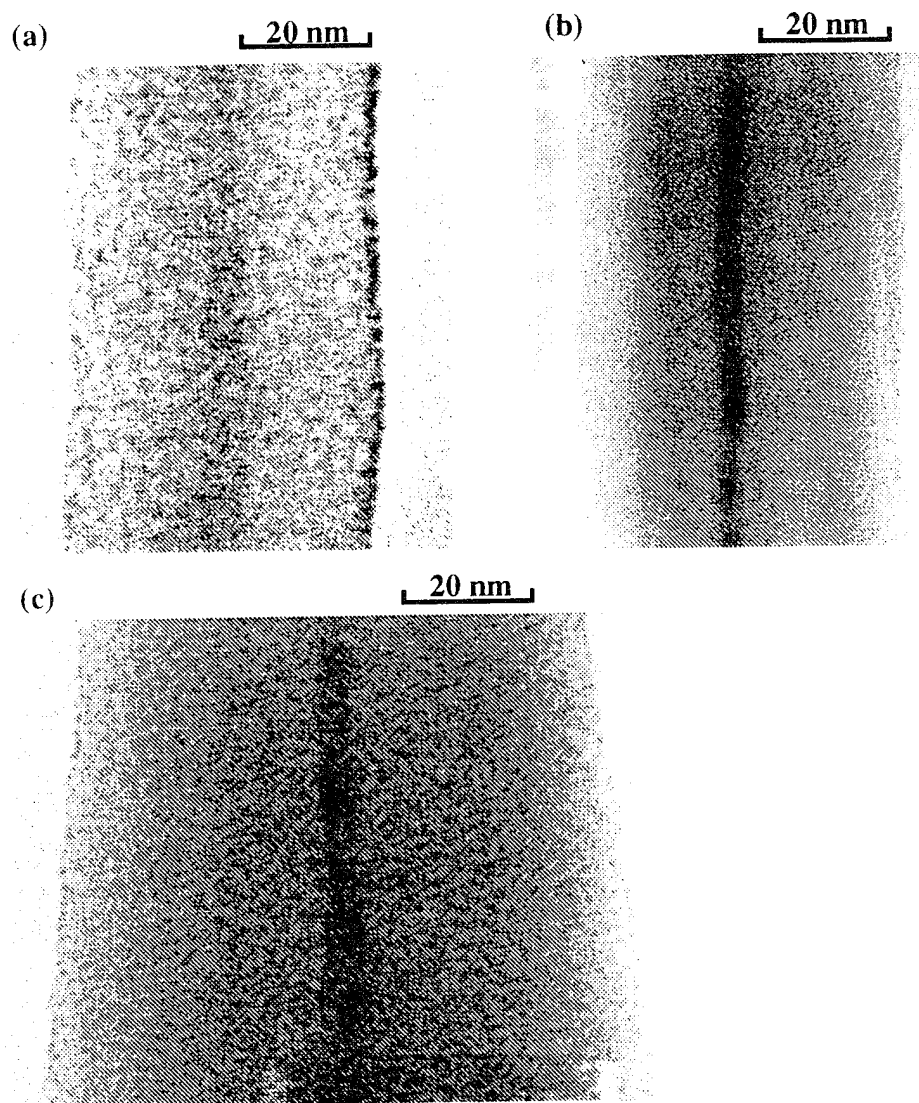


Figure 6. TEM micrographs of 5 nm silicon nano-wires. The 5 nm wide silicon nano-wires are produced at different dry oxidation temperatures: (a) 850°C for 16 hours ($d_i = 30$ nm, starting pillar diameter), (b) 900°C for 10 hours ($d_i = 40$ nm) and (c) 950°C for 10 hours ($d_i = 50$ nm). These are predicted by the viscous flow oxidation model.

Moreover, a polishing technique was perfected to yield a structure with silicon nano-wire array embedded in a metal layer (Fig. 7). This configuration is ideal for eliminating the background signal from bulk silicon substrate when performing optical characterization of nano-wires. In fact, with this structure, we have observed unique Raman spectra due to phonon confinement. This type of structure can also be used to fabricate self-aligned silicon field emission tip array. The close proximity of the extraction gate electrode is expected to yield very low turn-on voltage (below 10 V). Currently we are working to realize this experimentally.

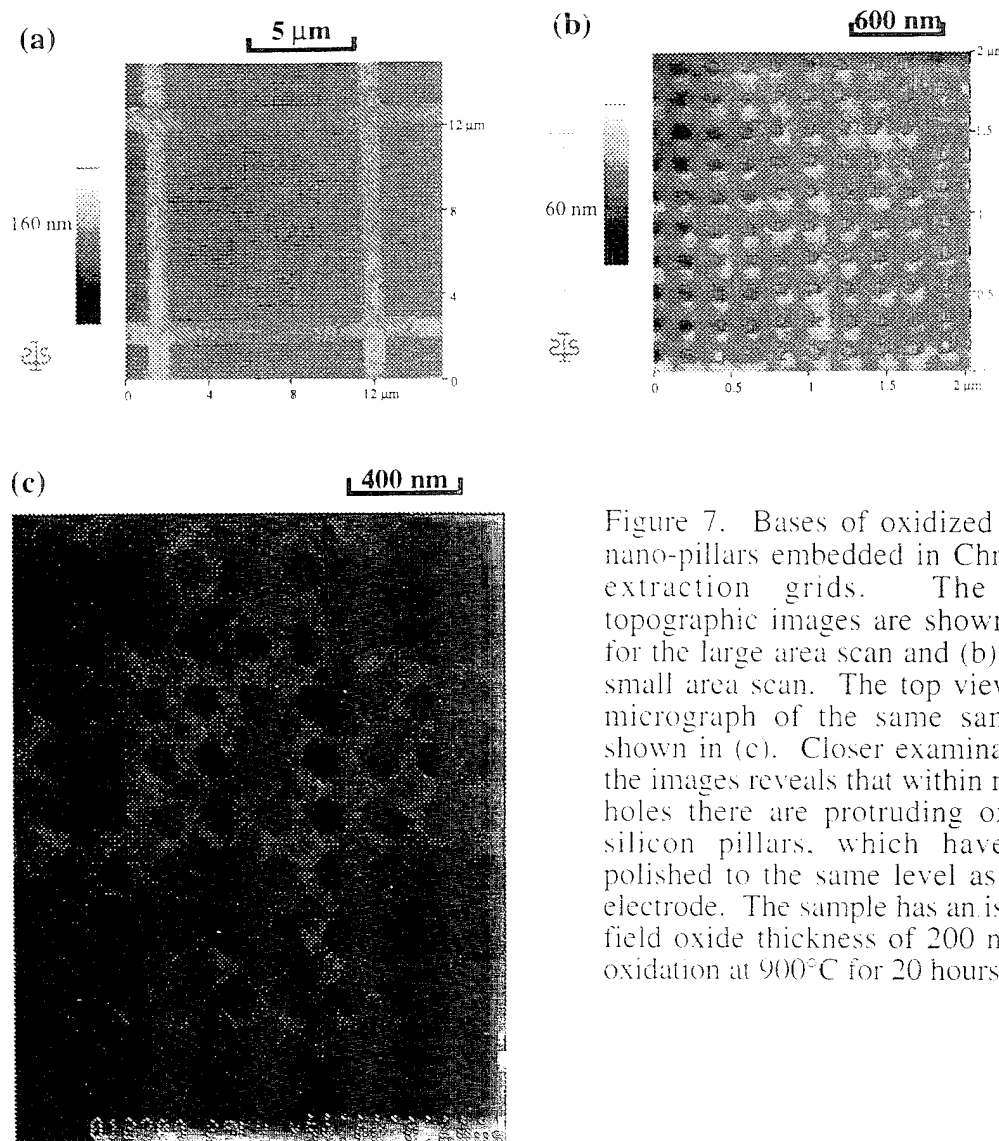


Figure 7. Bases of oxidized silicon nano-pillars embedded in Chromium extraction grids. The AFM topographic images are shown in (a) for the large area scan and (b) for the small area scan. The top view SEM micrograph of the same sample is shown in (c). Closer examination of the images reveals that within most Cr holes there are protruding oxidized silicon pillars, which have been polished to the same level as the Cr electrode. The sample has an isolation field oxide thickness of 200 nm (dry oxidation at 900°C for 20 hours).

References

- [Brown] W. F. Brown, Jr., *Ann. N. Y. Acad. Sci.*, **Vol. 147**, Art. 12, pp. 461-488, June 1969.
- [Canham] L. T. Canham, "Silicon quantum wire array fabrication by electrochemical and chemical dissolution of wafers", *Appl. Phys. Lett.*, **57** (10), 1046-8 (1990).
- [Liu] H. I. Liu, N. I. Maluf, R. F. W. Pease, D. K. Biegelsen, N. M. Johnson and F. A. Ponce, "Oxidation of Sub-50 nm Si Nanostructures for Light Emission Study", *J. Vac. Sci. Technol.*, **B, 10** (6), 2846 (1992).
- [Liu] H. I. Liu, D. K. Biegelsen, N. M. Johnson, F. A. Ponce, N. I. Maluf and R. F. W. Pease, "Silicon Quantum Wires: Oxidation and Transport Studies", *MRS 1992 Fall Proceedings*, Symposium F (1992).
- [Liu] H. I. Liu, D. K. Biegelsen, N. M. Johnson, F. A. Ponce, N. I. Maluf and R. F. W. Pease, "Fabrication of Si Nanostructures for Light Emission Study", *NATO ASI Series* **244**, 23 (1993).

- [Liu] H. I. Liu, D. K. Biegelsen, F. A. Ponce, N. M. Johnson and R. F. W. Pease, "Self-limiting oxidation of Si nanowires", *J. Vac. Sci. Technol. B*, **11** (6), 2532 (1993).
- [Liu] H. I. Liu, D. K. Biegelsen, F. A. Ponce, N. M. Johnson and R. F. W. Pease, "Self-limiting oxidation for fabricating sub-5 nm silicon nanowires", *Appl. Phys. Lett.* **64** (11), 1383 (1994).
- [New (a)] R. M. H. New, R. F. W. Pease and R. L. White, *J. Vac. Sci. Technol. B*, **Vol. 12**, No. 6, 3196, Nov/Dec 1994
- [New (b)] R. M. H. New, R. F. W. Pease and R. L. White, *J. Vac. Sci. Technol. A*, May/June 1995.
- [New (c)] R. M. H. New, R. F. W. Pease and R. L. White, submitted to IEEE International Magnetism Conference, April 1995.
- [Okada] R. Okada and S. Iijima, "Oxidation property of silicon small particles", *Appl. Phys. Lett.* **58** (15), 1662-3 (1991).
- [Tong] H. C. Tong, R. Ferrier, P. Chang, J. Tzeng and K. L. Parker, *IEEE Trans. Mag.*, **Vol. 20**, No. 5, 1831 (1984).

Unit: 3

TITLE : Investigation of a Metal Source and Drain Field Emission Transistor

PRINCIPAL INVESTIGATOR : C. R. Helms

GRADUATE STUDENT : J. P. Snyder

Scientific Objectives

The purpose of this work is to investigate various aspects of metal source and drain Metal Oxide Semiconductor Field Effect Transistors (MOSFETs). These aspects include ease of fabrication, performance over temperature, and scalability down to 0.1 μm .

Results to Date

Metal source and drain Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) were first investigated in the late 1960s [Lepse], and were thought to have certain advantages over their conventional (diffused source and drain) counterparts including a simplified process, the ability to make very shallow source and drain regions, low source and drain sheet resistance, and complete immunity to latch-up and parasitic bipolar effects. They proved to be poor performers however when compared to a similarly sized conventional MOSFET. The lower drive current in the 'on' state was attributed to the presence of a finite 'gap' between the edge of the poly gate and the edge of the platinum silicide (PtSi) source metal. The much higher leakage currents in the 'off' state originate at the drain end of the device, where electric fields cause the thermally assisted field emission of electrons from the drain into the silicon [Lepse] [Oh] [Koenke] [Sugino] [Tsui].

Until recently, the low temperature characteristics of these devices have not been investigated. The only exception to this is a 1968 paper [Lepse] in which 77 K I-V curves are shown and briefly discussed. Their device was fabricated with a non-self aligned, chemical vapor deposition (CVD) gate oxide process. The data shows a significant *decrease* in current drive at 77 K compared to room temperature.

Since 1993, several papers [Tucker] [Hareland] have reported on simulations on these and similar devices, and have shown acceptable drive current and short channel effects in devices with channel lengths down to 0.025 μm . The scalability of these metal source and drain devices is particularly impressive at low temperatures (77 K), as described by [Tucker]. It seems possible in light of these recent studies to build a metal source and drain device that has all the advantages previously mentioned, as well as superior scalability to well below 0.1 μm and free of the low drive and high leakage current problems. The only requirement is low temperature operation.

We report the first detailed experimental investigation of the low temperature, field emission characteristics of PtSi source and drain MOSFETs. I-V curves have been measured at various temperatures down to 4.2 K and for channel lengths down to 1 μm . Device fabrication has been optimized so that it is free from the 'gap' at the poly edge described earlier. As will be discussed, we observe a definite transition in the current flow mechanism of the device, from thermal to field emission, as the temperature is reduced below 100 K. In this low temperature 'field emission mode', the drive current when the device is 'on' is comparable to that of a conventional MOSFET, and short channel effects are not observable down to 1 μm , despite the fact that the substrate is nominally undoped.

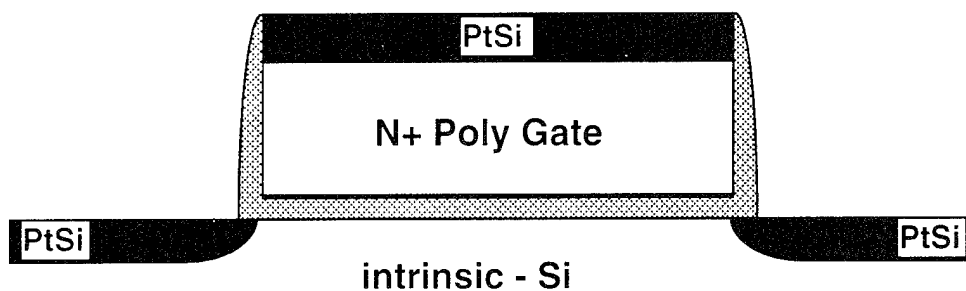
We started with <100> n type, nominally 3000 Ohm-cm ($1 \times 10^{12} / \text{cm}^3$) Phosphorous doped Si wafers. After a standard local oxidation of silicon (LOCOS) isolation process, 14.3 nm of gate oxide was thermally grown at 900 C in dry oxygen. Immediately following gate oxidation, ~120 nm of in-situ phosphorous doped amorphous Si was deposited by CVD at 580 C. This was followed by a silicon nitride cap of ~10 nm.

After anisotropic gate etching, and thin thermal oxide sidewall formation, the source and drain silicon and the top of the poly are exposed by a combination of dry and wet etching. About 15 nm of platinum is then sputtered on and silicided at 450 C for 90 minutes. A hot aqua regia dip removes the unreacted platinum from the sidewalls and the field oxide region. A standard back end process of contact hole formation, metal deposition and etch finishes the process.

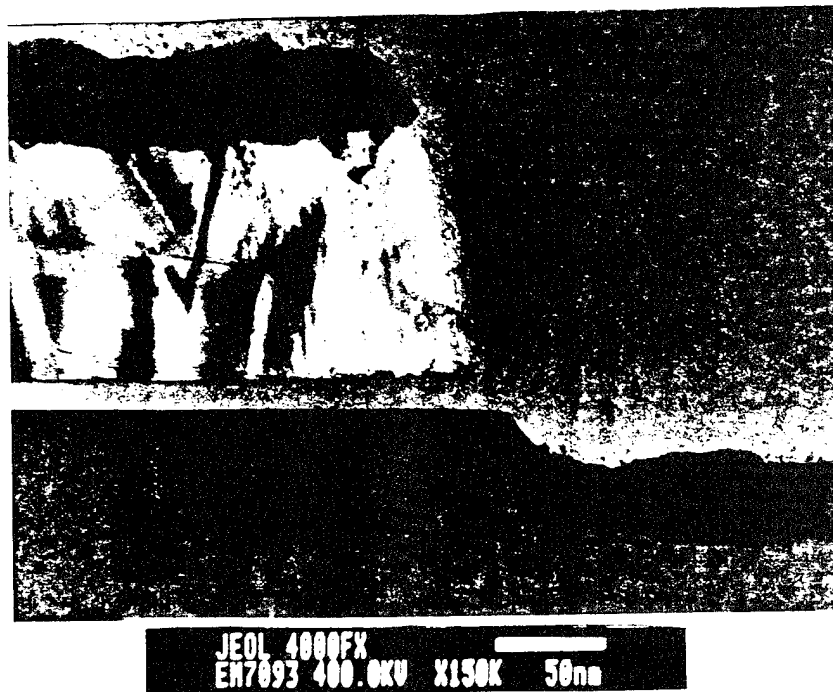
In Fig. 1, it is clear that the poly sidewall has not been silicided, and that the 'gap', described earlier, is virtually non-existent. Furthermore, one will notice small black dots on the sidewall oxide. These are presumably traces of platinum that were not removed by the hot aqua regia dip. Conveniently, they serve to mark the boundary between the thermal sidewall oxide and the surrounding CVD oxide.

The band diagrams of Fig. 2 demonstrate the operating principle of the device described in figure 1 at an intermediate temperature (~ 150 K) such that the various current flow mechanisms are observable. The band diagrams are drawn along a line from source to drain, just underneath the gate oxide, and show the Fermi levels of the source and drain PtSi, as well as the conduction and valence bands of the silicon substrate.

In Fig. 2(a), when the device is in its 'off' state with bias applied only to the drain, hole leakage current enters the channel by thermal emission over the sum of the 0.2 eV Schottky barrier and an electrostatic barrier present because of the difference in workfunction between the n+ poly



(a)



(b)

Figure 1. (a) Device structure and (b) Crosssectional Transmission Electron Micrograph of a successfully fabricated PtSi source and drain device with tox of ~ 14 nm. The source and drain regions as well as the top of the polysilicon have been silicided. Pt 'dots' on the polysilicon sidewall conveniently mark the border between the thermal oxide sidewall and the surrounding CVD oxide.

gate and the PtSi source. In this domain of gate voltage (V_g), the thermal emission regime, holes flow by diffusion from source to drain and the silicon bands in the channel are flat. Changing the gate voltage simply changes the amount of hole thermal emission current entering the channel, as is seen in the 'thermal emission characteristic' drawn in the plot of source current (I_s) vs. V_g . There is also the possibility of electrons being field emitted from the drain because of the high electric fields there, but this component of current does not show up in our measurements of source current and will not be discussed in this report.

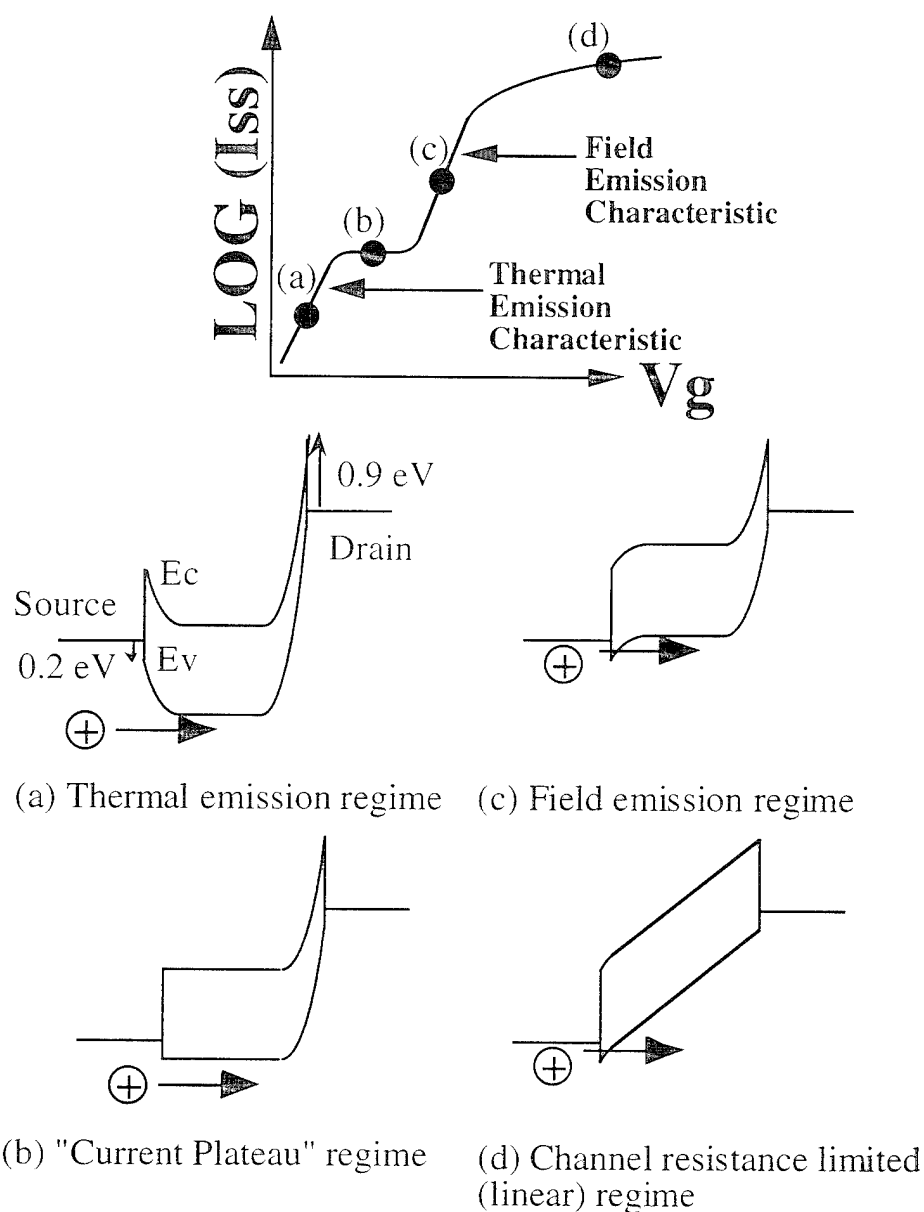


Figure 2. A band diagram description of the different current flow regimes seen in a typical source current vs. gate voltage plot. (a) Thermal emission regime (b) "current plateau" regime (c) field emission regime and (d) channel resistance limited regime.

Eventually, with increasingly negative gate bias, only the fixed Schottky part of the barrier to holes remains and the current is limited by thermal emission over this barrier [Fig. 2(b)]. In this 'current plateau' regime further increases in the magnitude of the gate voltage cease to have an exponential effect on I_s . The hole current is, for the most part, dependent only on the temperature and the barrier height (~ 0.2 eV), as is drawn in the topmost plot.

With high enough gate bias, holes eventually can be made to tunnel through the Schottky barrier and I_s once again begins to increase in an exponential fashion, this time along a 'field emission characteristic' [Fig. 2(c)]. The current is not yet large enough to give the silicon bands in the channel appreciable slope, which is to say that the current is still field emission limited and still travels by diffusion from source to drain, and is not yet channel resistance limited.

Finally I_s becomes large enough that the channel resistance begins to dominate and the holes travel by drift [Fig. 2(d)]. In this regime of V_g the current drive of the device is similar to that of a conventional MOSFET as the Schottky barrier has been rendered all but transparent to the flow of holes.

Drain curves (I_s vs. drain voltage (V_d)) and gate curves (I_s vs. V_g) were measured with a computer controlled HP 4140B DC voltage source/pA meter. A Lakeshore cryogenic probe station was used to perform measurements down to 4.2 K.

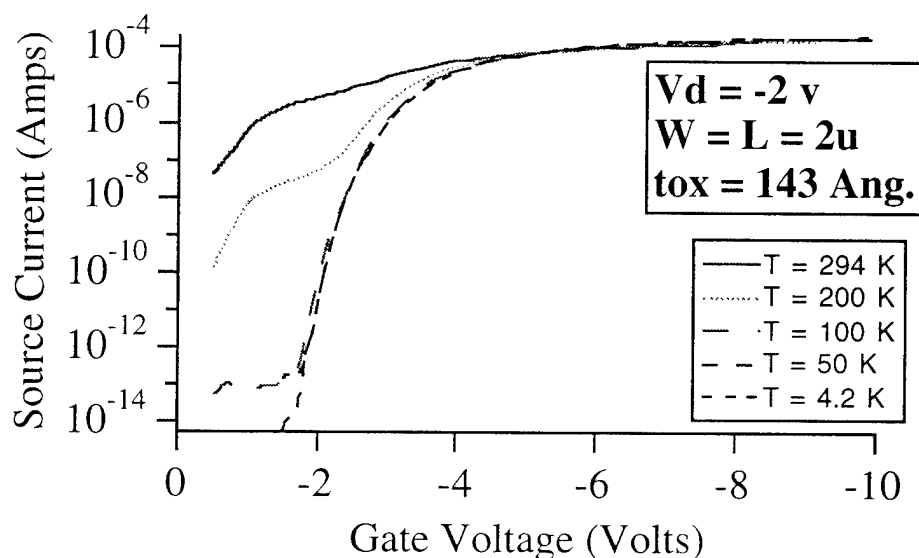
Figure 3(a) shows the gate curves of the device described in Figs. 1 and 2 with width=length= $2\text{ }\mu\text{m}$. Here the previously described thermal emission, plateau, field emission and channel resistance limited regimes are clearly seen, especially for the 200 K curve. As was mentioned previously, the plateau current is solely a function of temperature and barrier height and this dependence is observable. The plateau current drops exponentially with temperature, so that for temperatures less than about 100 K, all significant current flow (> 0.1 pA) occurs by the process of field emission and the device is being operated in the 'field emission mode'. It can be seen that this field emission characteristic is largely independent of temperature. Because n+ poly is used for the gate material, V_g must be brought to about -2 Volts before significant current begins to flow. Referring back to Fig. 2, this implies that even the band diagram in Fig. 2(c) could be used as an effective 'off' state. This could be realized for example, if p+ poly were used for the gate. The Schottky barrier alone is responsible for preventing the flow of current into the channel and thus it is clear why substrate doping is not required.

It is also possible to back out the effective PtSi - Si barrier height to holes from the thermal emission formula $I = AA*T^2\text{Exp}(q\phi_b/kT/(kT/q))$ using the plateau currents and corresponding

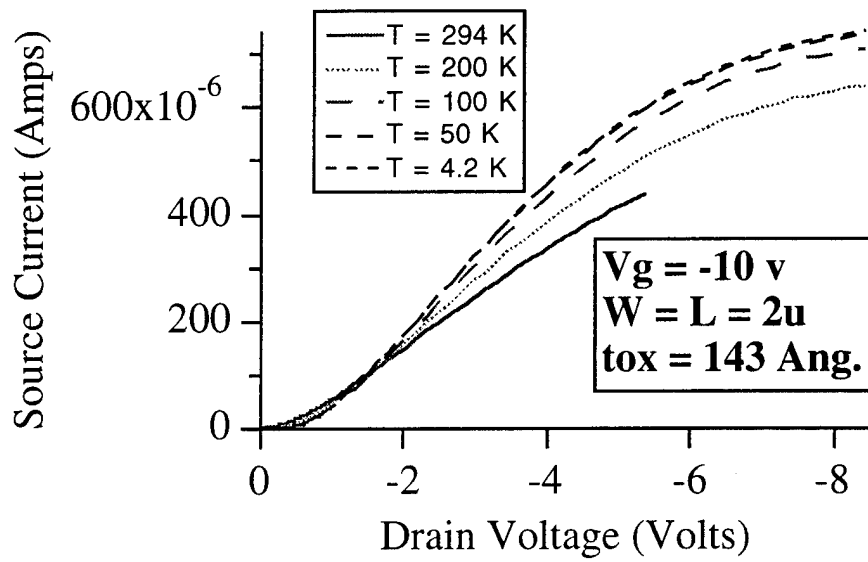
temperatures. This formula gives a barrier of ~ 0.195 eV, in very good agreement with published barrier heights of the PtSi - Si system [Mooney] [Weeks].

Figure 3(b) shows the drain curves in the linear region of the same device in Fig. 3(a). Current drive is comparable with that of a conventional MOSFET at the same temperature, according to 2-D simulations. It is also clear that the hole mobility and therefore the current drive improve with temperature, which contradicts previous data [Lepselter]. These two facts about the current drive are consistent with the Schottky barrier becoming transparent to the flow of holes and the current being limited solely by the channel resistance when the device is strongly 'on'.

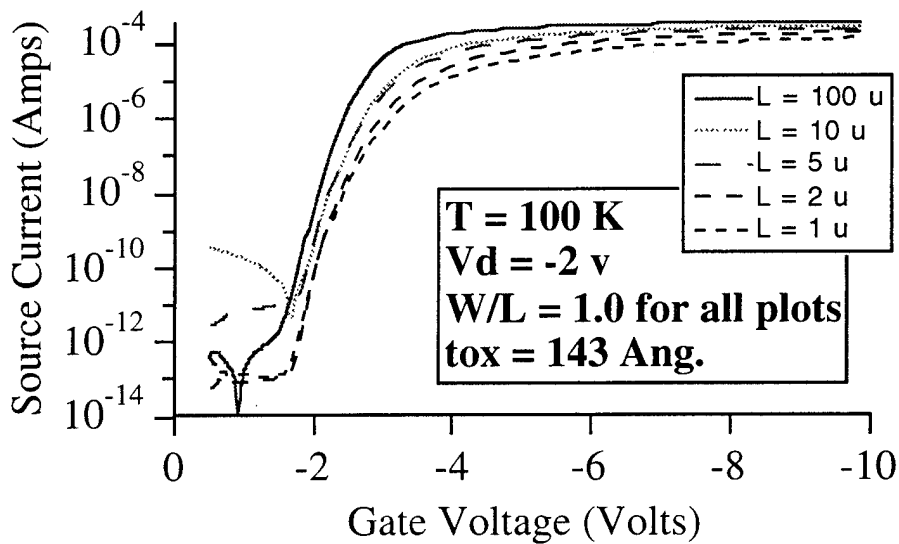
Finally, Fig. 3(c) shows short channel effects for devices fabricated, it will be remembered, on a nominally undoped substrate with channel lengths down to $1.0 \mu\text{m}$. Ignoring some spurious small V_g leakage current for the $10 \mu\text{m}$ device, it is clear that even the $1 \mu\text{m}$ device is well behaved and that no significant current flows until $V_g = \sim -2$ Volts, in agreement with [Tucker's] predictions of good scalability.



(a)



(b)



(c)

Figure 3. Experimental data. (a) Gate curves showing the different current flow regimes as described in Fig. 2. (b) drain curves for the same device as in (a). Drive current is comparable to a conventional MOSFET at the same temperature. (c) Short channel effects. Ignoring some spurious small- V_g leakage current on the $L=10\text{ }\mu\text{m}$ curve, there are no apparent short channel effects down to $L=1\text{ }\mu\text{m}$, despite the use of a nominally undoped substrate.

Work in Progress

We have successfully completed the fabrication of short channel devices ($L \sim 0.1 \mu\text{m}$) and are about to take low temperature measurements. We expect to see well behaved short channel effects. We will also investigate the previously discussed field emission of electrons from the drain when the device is in its 'off' state. Specifically, the temperature dependence of this component of leakage current will be studied. In addition to this, quantum effects (such as transconductance oscillations) which become apparent at 4.2 K will also be examined.

Possible Future Directions

First of all, an n-type device should be investigated. This can be accomplished by using Erbium Silicide (ErSi_2) as the source/drain metal, and an otherwise identical process flow used for the p-type (PtSi) device. Once this has been done, the next obvious step is to integrate the two devices on the same chip and develop a CMOS process. Direct write e-beam could be used to produce demonstration circuits with minimum feature sizes significantly less than $0.1 \mu\text{m}$. Such a circuit would be expected to set new performance standards for Si ULSI chips.

References

- [Hareland] S. A. Hareland, A. F. Tasch, C. M. Maziar, *Electronics Letters*, **29**, 1894 (1993).
- [Hareland] S. A. Hareland, A. F. Tasch, C. M. Maziar, *Proceedings of the 21st International Symposium on Compound Semiconductors*, September 18-22, San Diego, CA (1994).
- [Koenekke] C. J. Koenekke, S. M. Sze, R. M. Levin, E. Kinsbron, 1981 IEDM, 367.
- [Lepselter] M. P. Lepselter, S. M. Sze, *Proceedings of the IEEE*, 1400 (1968).
- [Mooney] J. M. Mooney, J. Silverman, M. M. Weeks, SPIE, *Infrared Sensors and Sensor Fusion*, **782**, 99 (1987).
- [Oh] C. S. Oh, Y. H. Koh, C. K. Kim, 1984 IEDM, 609.
- [Sugino] M. Sugino, L.A. Akers, M.E. Rebeschini, 1982 IEDM, 462.
- [Tsui] B. Tsui, M. Chen, *J. Electrochem. Soc.*, **136**, 1456 (1989).
- [Tucker] J. R. Tucker, C. Wang, J. W. Lyding, T. C. Shen, G. C. Abeln, 1994 SSDM, 322.
- [Tucker] J.R. Tucker, C. Wang, P.S. Carney, *Appl. Phys. Lett.*, **65**, 618 (1994).

[Weeks] M. M. Weeks, P. W. Pellegrini, SPIE, Test and Evaluation of Infrared Detectors and Arrays, **1108**, 31 (1989).

Unit: 4

**TITLE: The Electronic Structure and Interfacial
Properties of High Temperature Superconductors**

PRINCIPAL INVESTIGATOR: W. E. Spicer

JUNIOR FACULTY: Z.-X. Shen

GRADUATE STUDENTS: D. S. Dessau and D. S. Marshall

Scientific Objective:

The objective of this project is two-fold: (1) to gain a sound scientific understanding of the bulk electronic structure of high temperature superconductors (HTSCs), and (2) to apply this knowledge to the study of interfaces of the HTSCs with other technologically important materials.

Summary of Research:

(1) Normal State Fermi Surface of HTSCs

A question of central importance to understanding HTSCs is whether in their normal state, they are a Fermi liquid. This is important because these materials exhibit unusual physical properties that challenge the conventional wisdom of a Fermi liquid. The similarity of these materials to antiferromagnetic Mott insulators may be the reason for such behavior.

(a) Bi2212 thin film normal state spectra. In order to examine the changes in electronic structure as these materials are doped from their highest T_c , optimally doped state, to their antiferromagnetic Mott insulator state, a suitable material system was chosen. $\text{Bi}_2\text{Sr}_2\text{Ca}_{(1-x)}\text{Dy}_x\text{Cu}_2\text{O}_{8-\delta}$ (Dy doped Bi2212) was selected because it cleaves easily, its cleaved surface is stable in UHV conditions, and Dy when substitutionally replacing Ca dopes the material with electrons. A major technological hurdle had to be overcome for this study. The top-post cleaving technique had to be adapted to thin films grown by atomic layer-by-layer molecular beam epitaxy (ALL-MBE) [Marshall]. ALL-MBE had to be used because the Dy doping level and its substitutionality for Ca could only be controlled closely in this way. ALL-MBE also allows for higher substitution percentages than thermodynamically possible.

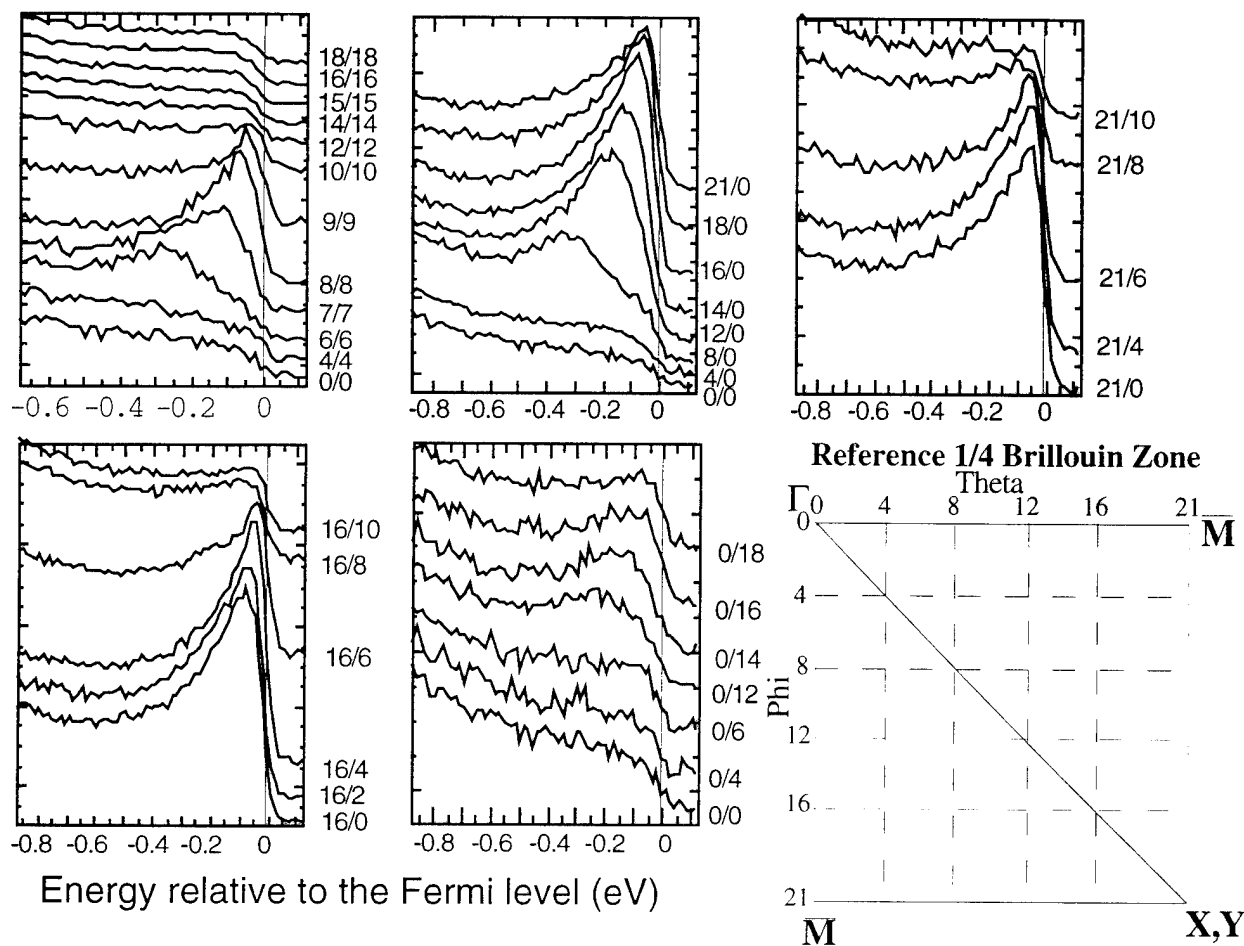


Figure 1: Angle resolved photoemission results from an optimally doped Bi2212 thin film grown by ALL-MBE. The θ, ϕ combinations on each spectrum correspond to brillouin zone locations according to the reference B-zone.

Studies of a fundamental nature have not been performed on thin films in the past because thin films have wide superconducting transitions (as measured by R vs. T) and a large number of systematic defects along the steps on the vicinal substrates. Not only did these defects raise questions as to the possibility of getting clean results from thin films, the defects cause bonding in the third dimension causing these otherwise micaceous materials to cleave much less easily.

The initial study of ALL-MBE grown material was confirmation that optimally doped thin films do in fact cleave and give results comparable to bulk crystals. Figure 1 shows several series of angle resolved photoemission spectra from an optimally doped thin film [Marshall]. This data is comparable to the best results from bulk crystals [Dessau]. These results are superior to any reported in the literature for thin films [Sakisaka]. The Fermi surface crossings were found for this thin film and compared to the Fermi surface of bulk crystals in Fig. 2 [Marshall, Dessau]. Excellent agreement is found between the results from thin films and bulk crystals.

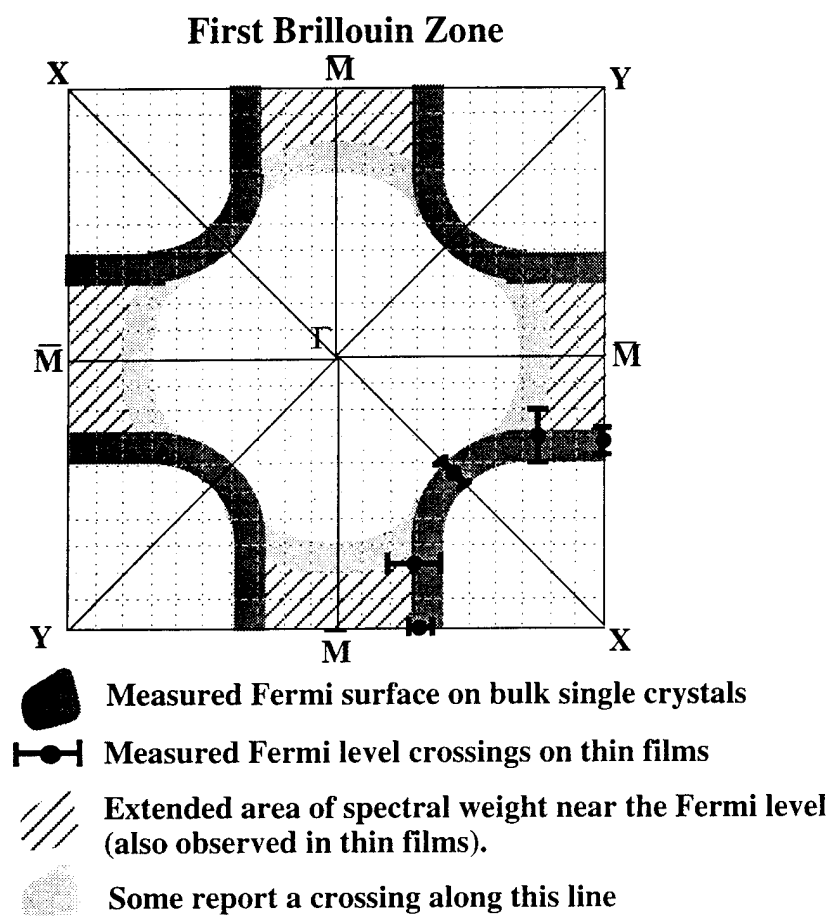


Figure 2: Comparison of thin film and bulk crystal Fermi surfaces.

(b) Bi2212 thin film superconducting spectra. The superconducting energy gap and dip at the \bar{M} symmetry point are some of the most sensitive photoemission features of these materials. Figure 3a shows that even these quite sensitive features are seen with thin films. This confirms that the basic electronic structure and low energy excitations in the films are very similar to those of bulk crystals, even though the films have higher defect densities and broader superconducting transitions. The anisotropy of the superconducting gap in k -space is observed by comparing Figs. 3a and 3b. In bulk samples, the gap is widest (as wide as 25 meV in the best bulk samples) at the \bar{M} symmetry point and has little or no gap at the G - X Fermi level crossing. The largest gap at \bar{M} in the thin films under study was 17 meV and is shown in Fig. 3a. The gap at the G - X Fermi level crossing (Fig. 3b) of this thin film is less than the experimental uncertainty of 4 meV. Such a large gap anisotropy is a unique feature of the high T_c superconductors.

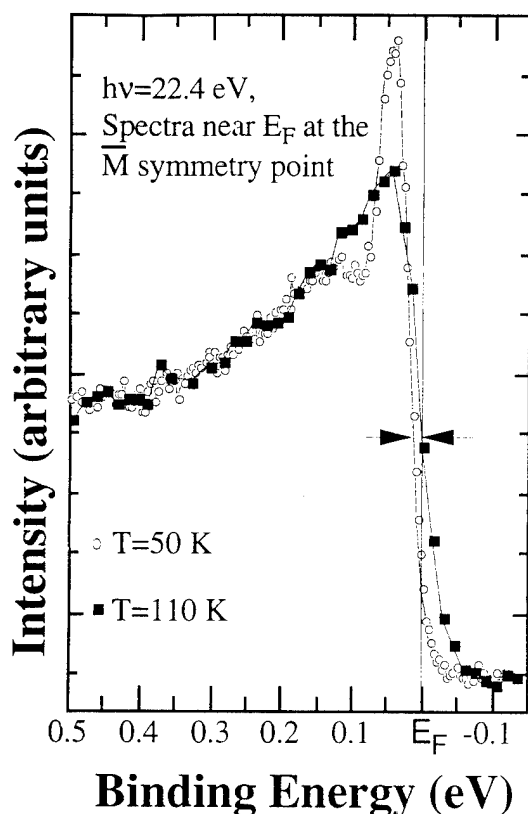


Figure 3a: Spectra from \bar{M} B- zone point taken above and below the superconducting critical temperature showing a gap of 17meV and the dip.

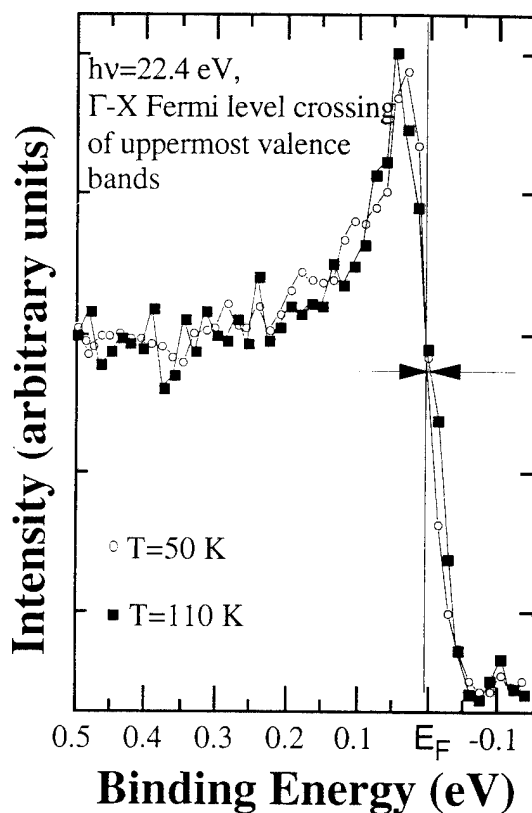


Figure 3b: Spectra at the G-X Fermi level crossing taken above and below the superconducting critical temperature showing little or no gap.

(2) Metal-Superconductor interfaces and the proximity effect

Our previous studies of metal overlayers concentrated on Ag and Au because of their low reactivity with the cuprates. A null result for the proximity effect was found with these materials. One explanation for why the proximity induced gap was not observed, was that the carrier density in these metals is several orders of magnitude larger than that of the cuprates. Our latest attempts included K doped C_{60} and Bi metal. Both of these materials have a low carrier density. Unfortunately, both materials also gave a null result when looking for the proximity effect. The reason for failure in both cases was reaction with the surface. We had hoped that the potassium would preferentially stay in the C_{60} lattice, but it did not, and reactions with the surface layer of Bi2212 resulted in electrical isolation of the C_{60} overlayer. In the case of Bi metal deposition, it was hoped that the BiO_2 surface atomic layer would be stable against the deposition of Bi metal on the surface. Bi core level studies showed that new oxidation states were forming upon deposition of Bi and this disruption of the top-most layer is believed to be responsible for the failure to see the proximity effect in this trial.

References

- [Marshall] D. S. Marshall, D. S. Dessau, D. M. King, C.-H. Park, A. Y. Matsuura, Z.-X. Shen, W. E. Spicer, J. N. Eckstein and I. Bozovic, preprint, submitted to *Phys Rev. B*, Dec. 1994.
- [Dessau] D. S. Dessau, Ph.D. Thesis, Stanford University, 1992.
- [Sakisaka] Y. Sakisaka, T. Komeda, T. Maruyama, M. Omchi, H. Kato, Y. Aiura, H. Yanashima, T. Terashima, Y. Bando, K. Iijima, K. Yamamoto and K Hirata, *Phys. Rev. B*, **39**, 9080, (1989).

JSEP Supported Publications

1. D. S. Marshall, D. S. Dessau, D. M. King, C.-H. Park, A. Y. Matsuura, Z.-X. Shen, W. E. Spicer, J. N. Eckstein and I. Bozovic, preprint, submitted to *Phys Rev. B*, Dec. 1994.

JSEP Related Publications

1. Z.-X. Shen, D. S. Dessau, B. O. Wells, D. M. King, W. E. Spicer, A. J. Arko, D. S. Marshall, L. W. Lombardo, A. Kapitulnik, P. Dickinson, S. Doniach, J. Dicarlo, T. Loeser, and C.H. Park, *Phys. Rev. Lett.*, **70**, 1553 (1993)
2. D. M. King, Z.-X. Shen, D. S. Dessau, W. E. Spicer, A. J. Arko, D. S. Marshall, J. Dicarlo, A. G. Loeser, C. H. Park, B. O. Wells, E. R. Ratner, J. L. Peng, Z. Y. Li and R. L. Greene, *Phys. Rev. Lett.*, **70**, 3159 (1993)
3. D. S. Dessau, Z.-X. Shen, D. M. King, D. S. Marshall, L. W. Lombardo, P. H. Dickinson, J. DiCarlo, C.-H. Park, A. G. Loeser, A. Kapitulnik and W. E. Spicer, *Phys. Rev. Lett.*, **71**, 2781 (1993)
4. E. R. Ratner, Z.-X. Shen, D. S. Dessau, B. O. Wells, D. S. Marshall, D. M. King, W. E. Spicer, J. L. Peng, Z. Y. Li and R. L. Greene, *Phys. Rev. B*, **48**, 10482 (1993)
5. J. Wu, Z.-X. Shen, R. Cao, D. S. Dessau, X. Yang, P. Pianetta, D. S. Marshall, B. O. Wells, D. King, J. Terry, D. Elloway, H. R. Wendt, C. A. Brown, H. Hunziker and M. S. de Vries, NOVEL FORMS OF CARBON SYMPOSIUM. Held: San Francisco, CA, USA, 27 April-1 May, 1992. (USA: Mater. Res. Soc, 1992. p. 235-40) (Conference paper - English)
6. D. S. Dessau, Z.-X. Shen and D. S. Marshall, *Phys. Rev. Lett.*, **71**, 4278 (1993).
7. D. M. King, Z.-X. Shen, D. S. Dessau, D. S. Marshall, C. H. Park, W. E. Spicer, J. L. Peng, Z. Y. Li and R. L. Greene, *Phys. Rev. Lett.*, **73**, 3298 (1994).

Unit: 5

**TITLE: On-Chip Thin Film Solid State
Micro-Battery**

PRINCIPAL INVESTIGATOR: S. S. Wong

GRADUATE STUDENT: J. Leung

Scientific Objectives

The objective of this work is to develop the fabrication technology and characterize the performance of thin film solid state micro-batteries that are suitable for monolithic integration with semiconductors.

Summary of Research

Recent advancements in the integrated circuit (IC) industry have resulted in devices with a reduced level of power consumption that are more suitable for portable electronics. Development in rechargeable battery technology has greatly improved the available energy density, lifetime and number of charging cycles. To further reduce the weight and size, thick film solid state batteries in the form of lithium coin cells have been marketed commercially [Akridge]. IC products that contain such batteries in the packages to provide non-volatile storage and continuous operation are available. The next advancement would be to integrate a solid state micro-battery onto the IC. Such a combination will offer several advantages:

1. The micro-batteries can directly replace bulk batteries in portable electronic systems. In addition to the reduction in weight and size, each IC will have its own battery, and hence the total charge capacity of the system could be much higher than that available from a single set of batteries.
2. The micro-battery can provide a backup energy source for non-volatile storage and continuous operation.

3. The micro-battery is more effective than a de-coupling capacitor in regulating the power supply level on the chip. A stable supply level is especially critical for low voltage and power applications.

The structure of a typical solid state battery is depicted in Fig. 1. In the thick film form, the various layers are sequentially laminated together. The composite structure is then cut into appropriate shapes and packaged into sealed containers. Recently, an experimental prototype of a solid state battery fabricated with thin film deposition techniques similar to those commonly used in the IC industry has been demonstrated [Jones]. This approach, in principle, is more appropriate for on-chip integration. Unfortunately, some of the materials used in thin film solid state batteries are unusual for ICs. The main focus of this work is to examine the compatibility issues, develop proper solutions, and demonstrate the monolithic integration of a micro-battery with an IC and evaluate the performance.

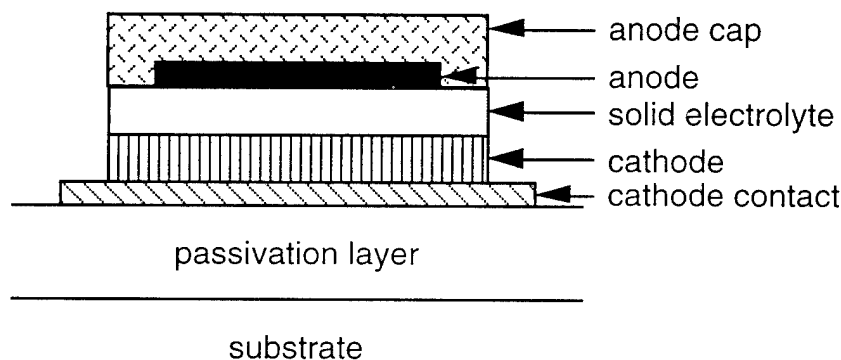


Figure 1. Solid state battery.

A prototype micro-battery has been fabricated on a silicon wafer. Various passivation layers have been evaluated. The most appropriate one is plasma enhanced chemical vapor deposited (PECVD) silicon oxynitride layer. This layer is impervious to the diffusion of lithium as confirmed by the results of SIMS analysis illustrated in Fig. 2. The cathode contact is evaporated chromium, which adheres well to oxide and is already widely accepted in the IC industry for photolithography mask and as a barrier metal for solder bump. The cathode is TiS_2 , which is sputtered from a composite target and is commonly used as a rechargeable electrode in thick film solid state batteries. The solid electrolyte is sputtered $6\text{LiI}-4\text{Li}_3\text{PO}_4-\text{P}_2\text{S}_5$. This is the most critical layer for the structure and various other options will be studied. The anode is evaporated lithium, and no anode cap is used in this first experiment. The sample therefore has to be stored and tested in an argon ambient

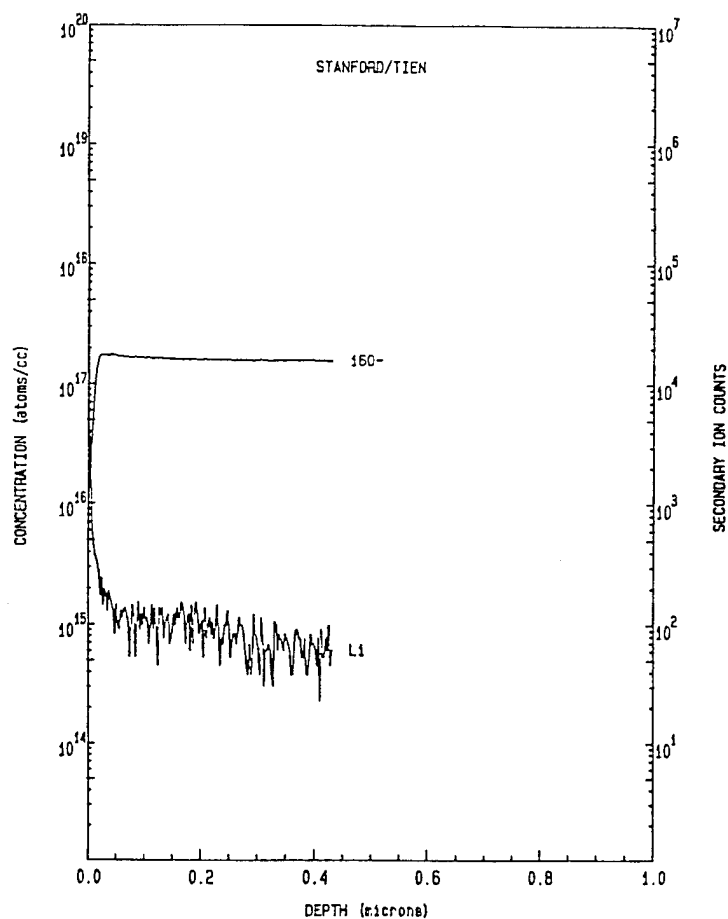


Figure 2. SIMS analysis of Li/oxynitride/Si sample after heat treatment at 100 C. Li and oxynitride was removed prior to SIMS.

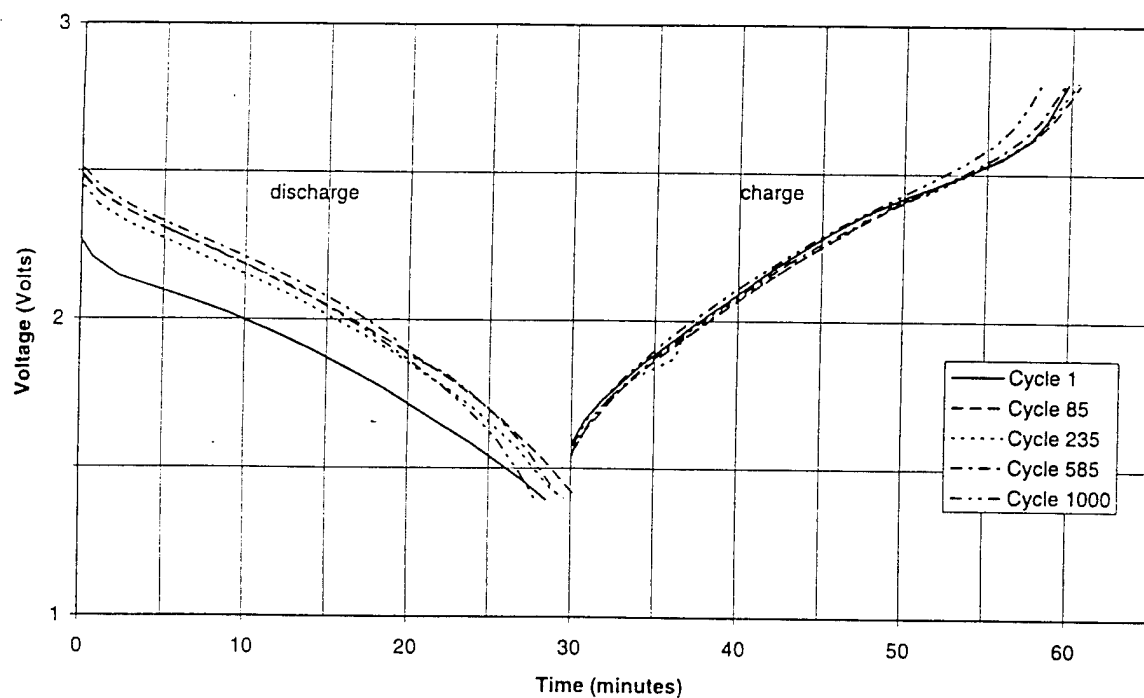


Figure 3. Charge and discharge behavior of micro-battery. The current levels are about 50 mA.

to prevent oxidation. Figure 3 shows that charge and discharge behavior of the micro-battery is quite stable even after 1000 cycles.

In the next phase of this project, we plan to investigate the performance behavior of various spun-on polymeric solid electrolyte, and the integration of the micro-battery with BiCMOS circuitry.

References

- [Akridge] J. Akridge and H. Vourlis, *Solid State Ionics*, 18-19, p. 1082, 1986.
- [Jones] S. Jones and J. Akridge, *Solid State Ionics*, 13-56, p. 628, 1992.

Unit: 6

**TITLE: CVD Epitaxial Germanium *n*-Channel FETs formed on
Si substrates using strain-relief layers**

PRINCIPAL INVESTIGATOR: K. Saraswat

GRADUATE STUDENT: D. Connelly

Abstract

Different epitaxial strain-relief techniques are investigated to yield Germanium-channel field-effect transistors of different degrees of surface-film strain integrated on silicon substrates. The effect of surface strain on transport parameters is to be used to investigate the transistion from a *L* to *X* conduction band minimum in the germanium. Also to be investigated is the possibility of *n*-channel transistors in high-germanium germanium-silicon alloys. Integration with a silicon process will be addressed.

Scientific Objectives

The following are the primary objectives of this project:

- To fabricate *n*-type Ge-channel MOSFETs on a Si substrate.
- To investigate the effect of different degrees of compressive strain on the electron transport properties in germanium inversion layers.
- To compare different schemes for the formation of strain-relief structure formation including blanket graded epitaxy, selective graded epitaxy, and graded epitaxy on ultra-thin silicon-on-insulator.
- To assess the utility of high-germanium content *n*-channel MODFETs in high-speed transistor applications.

Prior Art

The development of strained layer epitaxy of GeSi alloys on silicon substrates sparked interest in the development of heterostructure devices using silicon-based technology. Much of the work can be characterized in one of two categories, vertical heterostructure bipolar transistors (see for example [King]), in which the primary interest is the band-gap difference between the base alloy and the emitter alloy, and confined-carrier field-effect devices (see for example [Pearsall86] and [Daembkes]) in which the parameter of interest is the conduction band offset (for n -channel devices) or the valence band offset (for p -channel devices).

The biaxial compressive strain formed when GeSi with non-zero x is deposited on silicon enhances the natural positive valence band offset of the GeSi relative to silicon [Pearsall89]. Representative is the work by the group from UCLA [Nayak] in which a 10 nm "undoped" unstrained silicon layer is deposited on an n -type Si substrate. An undoped 15 nm strained $\text{Si}_{0.80}\text{Ge}_{0.20}$ layer is then deposited to form the channel region. It is capped with a 10 nm silicon layer. A 5 nm SiO_2 layer is then thermally grown to form the insulator, consuming some of the underlying silicon. The structure is capped with a polycrystalline silicon gate electrode. The estimated 0.15 eV valence band discontinuity confines most of the holes to the $\text{Si}_{0.80}\text{Ge}_{0.20}$ layer for the initial portion of the superthreshold gate bias regime. The Princeton group [Garone] fabricated a similar structure with a 10 nm $\text{Si}_{0.60}\text{Ge}_{0.40}$ well capped by a 7.5 nm silicon spacer and a 10 nm gate oxide with an aluminum gate electrode.

For electron-confinement structures things are more complicated. For unstrained material of low-to-moderate germanium concentrations the conduction band consists of six degenerate ellipsoids aligned along the x , y , and z axes in momentum space. For material under [001]-directed strain the degeneracy is broken --- the z -directed ellipsoid is either raised or lowered in energy relative to the x and y valleys. Since the z valley exposes its carriers during conduction in the x - y plane to only the light transverse effective mass it is preferable to raise the energy and thereby reduce the carrier population of the x and y ellipsoids. This is done by depositing the channel material in biaxial tension.

Reported work to date has been on structures utilizing strained silicon as the channel material. Representative is the work from Stanford [Welser]. They had two forms of their device. One started with a relaxed {001} $\text{Si}_{0.70}\text{Ge}_{0.30}$ surface on which was grown a strained silicon layer. Subsequent oxidation of the silicon resulted in a 12.8 nm gate oxide over a 4.6 nm strained silicon channel well. The other used a $\text{Si}_{0.71}\text{Ge}_{0.29}$ surface on which was grown an 8.0

nm strained silicon layer covered with a 7.2 nm $\text{Si}_{0.71}\text{Ge}_{0.29}$ spacer and a "sacrificial" strained silicon cap. Thermal oxidation to form the 12.8 nm gate oxide fully consumed the cap. Thus the former devices had surface channels while the latter had buried channels for moderate superthreshold biases.

IBM has published results of Schottky-gated n -channel structures using both Molecular Beam Epitaxy and Ultra High Vacuum Chemical Vapor Deposition (UHVCVD) [Ismail] [Wang]. They used a starting relaxed surface with a 30% germanium content. Their channel was formed in a 10.6 nm strained silicon layer.

The key difficulty in the formation of these structures is the preparation of the initial surface. Ideally, if a surface of a given alloy composition is needed, a wafer of that composition should be used. Unfortunately wafers of arbitrary germanium content are not available --- silicon wafers are widely available while germanium wafers are available at considerably higher cost. A solution is to deposit a relaxed "buffer layer" in which threading dislocations are isolated below the surface to translate the surface composition to the desired value from that of the substrate. Leaders in this technique include AT&T with Molecular Beam Epitaxy and IBM with CVD and MBE. All reference cases described here begin with (100) silicon wafers.

Fitzgerald reports the results of linearly grading the germanium content from zero up to 53% using MBE at 900° C. The alloyed germanium content is ramped at 10% per micrometer. The high temperature is used to prevent the accumulation of stress in the films before relaxation occurs, increasing the number of threading dislocations and canceling the benefits of compositional grading. They fabricate ungated electron confinement structures [Xie] with good results. Other workers [Schaffler] showed that increasing the gradient to 45% per micrometer and decreasing the deposition temperature to 750° C can still yield significant advantages over abruptly stepped buffer layers.

IBM has generated strain-relief layers using both continuous grading as per AT&T (see [Legoues91] and grading in discretized steps [Meyerson] using both MBE and CVD. Tsang reports step-grading from pure silicon to pure germanium with fewer than 0.01 threading dislocations per square micrometer in the top germanium film. The deposition is done using UHVCVD with composition graded in 40 steps at approximately 20% per micrometer. The quality of the film is sensitive to the deposition temperature, with 450° C optimal for the pure-germanium portion. This is described by [Legoues92].

The Stanford group [Welser] used, for example, a graded layer of germanium composition from 6% to 30% continuously graded over 1.6 μm deposited via CVD using "Limited Reaction Processing" at 750° C. They had difficulty grading beyond 50% germanium starting from pure silicon using their technique, although IBM's positive results show it can be done.

Proposed Work

Single-crystal GeSi alloy exhibits a peak valence band energy which increases steadily with increasing Ge content. The energy of the sixfold-degenerate X (used here to signify all six $\langle 100 \rangle$ directions) conduction band valleys is relatively insensitive to the Ge content in unstrained material. Up to approximately 80 atomic-percent Ge these X -valleys have the lowest energy of the conduction states in the material. At higher Ge concentrations, however, the strong alloy-dependence of the eight-fold degenerate $\langle 111 \rangle$ L -valleys brings them to a lower energy.

Due to the dependence of the valence band energy on alloy content across the material spectrum most unipolar heterostructure devices built in the low-Ge regime have used holes as their carrier. n -type devices have been built, however, exploiting the strain-dependence of the conduction band minimum.

When (100) silicon is deposited pseudomorphically on a thick unstrained crystalline GeSi alloy the silicon is in biaxial tension, with decreased lattice spacing in the growth direction (z) and increased lattice spacing in the two orthogonal directions (x and y). The result is that electrons in the z -valleys ($[001]$ and $[00\bar{1}]$) exhibit a reduced energy relative to those in unstrained silicon while the x and y valleys see an increase in the energy of their states. (See [Pearsall89] for a good overview of the strain effects on GeSi bands.) The advantages are two-fold. First, since the unstrained GeSi substrate has similar conduction band energies to unstrained silicon, the Si now has a reduced conduction band energy relative to the surrounding material and electron confinement can be achieved. The second advantage is that these valleys exhibit a transverse effective mass lower than their longitudinal effective mass. Since conduction in the channel by z -valley electrons will be characterized by the lower transverse effective mass while electrons in the other four valleys will be subject to a mixture of the longitudinal and transverse effective masses, preferential occupation of the z valleys results in a decrease in net effective mass and a corresponding increase in mobility for appropriate carrier densities. The stress-induced electron confinement for devices in principle works for alloys from zero Ge up to

approximately 80 atomic percent Ge. However, work to date has focused on using strained silicon as the channel material.

In Ge-rich material there is therefore available two mechanisms to yield band offsets. If the unstrained starting material is (100) $\text{Si}_{0.25}\text{Ge}_{0.75}$ then application of a strained layer of pure Ge will result in a reduced conduction band energy due to the lower energy of the *L*-valleys (due to symmetry the effect of the [001] compression on the $\langle 111 \rangle$ *L*-valleys is small). Growth of a strained $\text{Si}_{0.50}\text{Ge}_{0.50}$ film on the same substrate will result in reduction of the *z*-valley energies relative to the unstrained material. These offsets could be used in the formation of confined-electron structures.

Of further interest in Ge channel devices is in which valleys the conduction band minimum occurs. As the degree of [001] compression is increased via a lowering of the effective substrate germanium content, the energy reduction of the *x* and *y* valleys increases the population of electrons occupying them until they become the principle repository for channel electrons. The effect of this transistion on electron mass and electron scattering is of significant importance.

Of practical interest is the formation of the relaxed buffer layer. Linear grades can be done via different temperature schedules to confine stress-relieving defects below the surface. These grades can be executed either on a blanket wafer or in regions defined in a surface oxide layer. Another option is the formation of a graded buffer layer on ultra-thin silicon-on-insulator, decreasing the energy needed to relax the surface.

The Past Year

Work has progressed in the development along three basic fronts. The first is the basic MOS process. This involved development of a process sequence compatible with the presence of high-germanium GeSi, design of the photomasks needed for process photolithography, and testing of the process on silicon wafers.

The second is the formation of strain-relaxation structures using the Stanford ASM reactor. Initial experiments were done using disilane and dilute germane as the source gasses for the silicon and germanium components of the film. To allow lower-temperature processing, a switch was made to silane as the silicon source. To allow more robust gas flows for higher germanium contents, a switch was made from dilute to pure germane, and a dual mass-flow-

controller control scheme was installed to allow broad control over germane flows. Films were grown using buffers of different lengths, different base temperatures, and different gas flow, pressure, and temperature profiles as a function of time. Films were assessed visually, via cross-sectional transmission electron microscopy, surface atomic force microscopy, Rutherford back scattering, and secondary ion mass spectroscopy. Finally, masks were made to generate arrays of diode and capacitor structures. Fabrication revealed the reverse bias leakage of arsenic-implanted diodes to be excellent, limited only by the poor quality of the simple CVD SiO₂ passivation used.

Associated with the second front is the investigation of alternate relaxation schemes. The first is selective graded epitaxy. The addition of HCl to the gas mix during the high-silicon-fraction portion of the growth yields selective deposition versus silicon dioxide. In the high-germanium regime, deposition is selective without the need for HCl. The use of selective epitaxy greatly simplifies device isolation and provides a convenient process for the integration of graded buffer devices with devices built in Si. Preliminary investigations have also been done on the use of ultra-thin silicon-on-insulator substrates as a starting point for relaxed buffer formation.

The third front is the development of the gate insulator. Previous work on FETs in pure germanium has shown that germanium oxinitrides yield satisfactory results [Hymes]. Experiments on these and other schemes involving deposited oxides have been tested using germanium wafer fragments. Tests will continue on germanium buffers deposited epitaxially on graded layers.

References

- [Daembkes] Daembkes, et al; *IEEE TED*, **33**:663 1986.
- [Fitzgerald] Fitzgerald, et al; *APL* **59**:811-813 1991.
- [Garone] Garone, et al; *IEEE EDL* **12**(5):230-232 1991.
- [Hymes] Hymes, et al; *J. of the Electrochem. Soc.*, **135**(4):961-965 1988.
- [Ismail] Ismail, et al; *IEEE EDL* **13**(5):229-231 1992.
- [King] King, et al; *IEEE EDL* **10**:52 1989.
- [LeGoues91] LeGoues, et al; *Phys Rev Letters*, **66**(22):2903-2906 1991.
- [LeGoues92] LeGoues, et al; *Journal of Applied Physics*, **71**:4230-4243 1992.
- [Meyerson] Meyerson, et al; *Applied Phys Letters*, **53**(25):2555-2557 1988.

- [Nayak] Nayak, et al; *IEEE EDL* **12**(4):154-156 1991.
- [Pearsall86] Pearsall and Bean, *IEEE EDL* **7**:308 1986.
- [Pearsall89] Pearsall, *CRC Critical Reviews of in Solid State and Materials Sciences*, **15**(6):551-600 1989.
- [Schaffler] Schaffler, et al; *Semiconductor Device Tech*, **7**:260-266 1992.
- [Tsang] Tsang, et al; *Appl Phys Let*, **62**(10):1146-1148 1993.
- [Welser] Welser, et al; *IEDM Tech Digest* 1000-1002 1992.
- [Wang] Wang, et al; *Materials Research Society Symp Proc*, **220**:403-408 1991.
- [Xie] Xie, et al; *Materials Research Society Symp Proc*, **220**:413-417 1991.

Unit: 7

TITLE: Portable Video on Demand in Wireless Communication

PRINCIPAL INVESTIGATOR: T. H. Meng

GRADUATE STUDENT: K. Precoda

1. Scientific Objectives

The provision of voice, video, text, and graphics data accessible by people unconnected to a physical interconnect medium has become the basic framework through which the information revolution evolves.

This research aims at providing portable digital video-on-demand in a wireless environment. The three main technological issues for portable video communication to be addressed are automatic bandwidth control, error recovery, and low-power implementation.

Applications of this technology to the military arena include wireless multimedia information exchange for soldiers on the field, video/graphics communication between forces on land and sea, from reconnaissance and targeting, to other aspects of a military mission.

2. Summary of Research

Development of error-resilient compression algorithm

An error-resilient compression algorithm exceeding the performance of standard image compression algorithms such as JPEG has been developed and finalized. This algorithm outperforms JPEG at all bit rates of interest, without requiring the use of variable-rate entropy codes. In addition to this high degree of compression efficiency, this compression algorithm embeds error-resiliency in the coding process, guaranteeing consistent video quality under all error conditions without the use of error-correcting codes. Its hardware complexity is less than half of that of the JPEG-like algorithms, allowing for a low-power implementation without the need of any external hardware support such as frame buffers and video control circuitry.

Design and fabrication of the video decoder chip set

We completed and fabricated a low-power video decoder chip set based on the algorithm mentioned above, which includes a subband decoder chip, a pyramid-vector-quantization (PVQ) decoder chip, and a low-power D/A converter (DAC) for the control of a color LCD display. The low-power video decoder chip set capable of decoding 30 frames/sec of video at a power level of 8 mW, 100 times below existing video decoding chips, was designed and tested. Both decoding chips have been tested and successfully worked together in tandem.

Development of scalable compression techniques

We have developed a psycho-visual based scalable compression algorithm that allows adaptive rate control to accommodate different bandwidths available in an open network. This algorithm will be applied to distributing Stanford televised lectures to the Internet, where video data will be transmitted and decoded in real time based on the instantaneous bandwidth available to each user on the net. The compression performance evaluation is based on our earlier work on psycho-visual distortion measure. We have also designed and implemented an end-to-end software only scalable video delivery system to demonstrate the effectiveness of this algorithm.

1. Error-Resilient Compression Algorithm

Because our portable video-on-demand system is to be embedded in a wireless communication environment, the first step in designing a suitable compression algorithm is to analyze the wireless channel characteristics. Wireless channels, depending on the channel modulation techniques used, display a wide range of error patterns. Experiments with mobile receivers and transmitters show fades exceeding 10 dB about 20% of the time, and at 15 dB (measuring limit) up to 10% of the time [Lee], causing bursty bit errors in the received data stream. Bursty bit errors are usually handled by interleaving the transmitted data stream over a certain period of time so that if consecutive bit errors occur (as is typical with bursty bit errors) in the received data stream, the error pattern appearing in the decoded data stream will resemble random bit errors.

To reduce the effects of channel random bit errors, error-correcting codes are usually used. With the knowledge of a target channel bit error rate (BER) and the channel's signal-to-noise ratio,

error-correcting codes can often be very effective. However, for wireless and mobile channels that experience very low BERs most of the time but with intermittent severe channel degradation occasionally, error-correcting codes may not be the best solution. On the one hand, under low distortion conditions (low BERs), error-correcting codes add unnecessary overhead to transmission bandwidth, which could have been used for transmitting compressed video data to improve video quality. On the other hand, under severe distortion conditions in which the BER exceeds the designed capacity, error-correcting codes may actually introduce more decoded errors than received ones. On top of all this, error-correcting codes require additional hardware at the decoder unit, making the design of our portable decoder a more difficult task.

Our approach to error resistance is to embed error-resiliency into the compression process so that if channel error does occur, the effects of it will be a gradual degradation of video quality and the best possible quality will be maintained at all BERs. Our compression algorithm utilizes lattice vector quantization (VQ), of which pyramid vector quantization (PVQ) is a special case, together with subband decomposition. Lattice VQ provides several distinct advantages. First, it is a fixed-rate code, which results in hardware simplicity and prevents catastrophic error propagation. Second, because of its regular lattice structure, lattice VQ allows for simple real-time decoding and encoding. Third, when optimized for the statistics of image data, lattice VQ provides excellent rate-distortion performance for moderate to high bit rates, achieving the compression performance of entropy codes asymptotically [Fisher] [Tseng].

Subband decomposition offers additional compression by decomposing the information energy of image data into several frequency bands. Unlike the DCT used in standard compression, the subband approach does not introduce blocking artifacts. Subband decompressed images are therefore usually considered to be more visually pleasing. In addition, the hierarchical nature of subband decomposition allows for flexibility in bit allocation, with different bit rates assigned to different subbands based on information content, visual importance, and subband size. Finally, subband decomposition provides many possible algorithm trade-offs to achieve a low-power implementation.

Figure 1 graphically shows the performance of our algorithm under various noisy channel conditions. Under most error conditions, our subband/PVQ algorithm performs better than JPEG with or without protection, as expected from the fact that our algorithm outperforms JPEG even without channel error. Over a wide range of low BERs, the subband/PVQ algorithm delivers images of higher quality because the whole channel bandwidth can be used for transmitting compressed video data. On the JPEG with error correction curves, we note that error-correcting

codes do an excellent job of eliminating bit errors, but start to fail at past BER of 10^{-2} . Once the BER increases past this point, catastrophic errors occur, causing severe block loss and rapid drop in peak signal-to-noise ratio (PSNR) performance. Our fixed-rate scheme, however, maintains a gradual decline in quality, even under severe BER conditions. This gradual degradation performance makes the subband/PVQ scheme well suited for situations where image quality must be maintained under deep channel fades and severe bit loss, a characteristic of wireless channels.

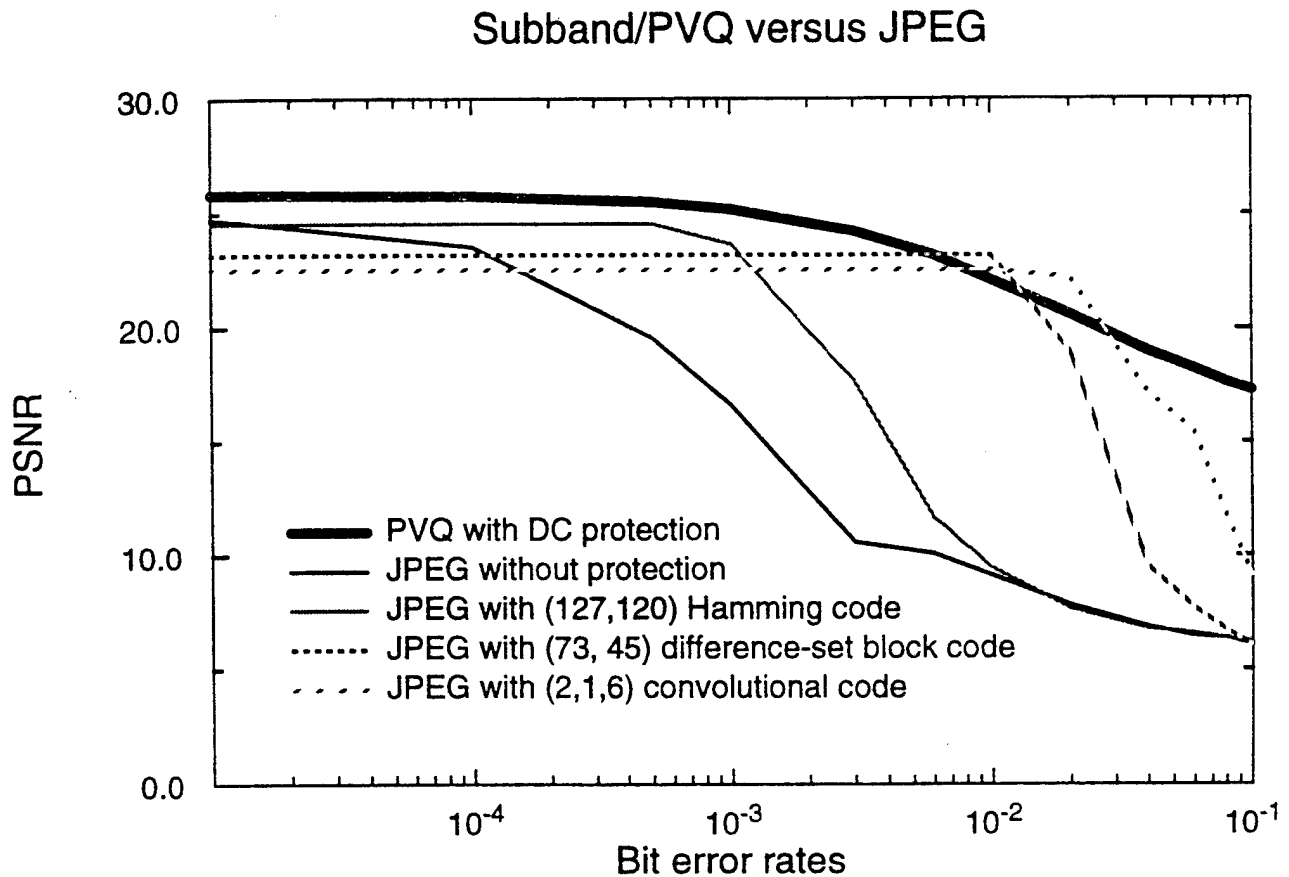


Figure 1. Peak signal-to-noise ratio vs. bit error rates for subband/PVQ and JPEG with error-correcting codes.

2. Portable Video-on-Demand Decoder Design

In comparison with the C-Cube JPEG decoder, implemented in 1.2 μm CMOS technology dissipating 2 W in decoding 30 frames of video per second [Purcell], our subband/PVQ decoder is more than 100 times more power efficient, not accounting the power dissipated in accessing off-chip memory necessary in the JPEG decoding operation. Within this factor of 100, a factor of 10 can be easily obtained by voltage scaling of the power supply. Reduced supply voltage, however, increases circuit delay. This increase in delay needs to be compensated for by duplicating hardware, or chip area, to maintain the same real-time throughput [Chand]. The fact that our decoder module consists of only two custom chips, without requiring any off-chip memory support, indicates the simplicity of our decoding operation, one of the established goals of our compression algorithm design. How we achieved the other factor of 10 reduction in power is one of the focuses of this project.

In designing the subband decoder, we emphasized a low-power implementation without introducing noticeable degradation in decompressed video quality. As memory accessing is by far the most power-consuming operation, the main design strategy has been to eliminate memory accesses in favor of on-chip computation. A two-dimensional subband decoder has been designed for real-time video decompression in low-power applications. The chip dissipates less than 1.2 mW at a 1 V supply, delivering subband decomposition at 1.3 Mpixels/sec, for display of 176 pixels wide, 240 lines, and 30 frames per second color video. The chip is capable of reconstructing 4 levels of hierarchical subband structures for images up to 256 pixels wide and requires no external hardware support such as frame buffers or video control.

Figure 2(a) illustrates the power dissipation at the maximum operating frequency for various supply voltages. Figure 2(b) illustrates the variation in energy and delay as a function of the supply voltage. At a supply voltage of 1 V, the decoder chip operates at the required 3.2 MHz real-time video rate and dissipates under 1.2 mW. From measurement, power consumption in the control section remains a small percentage despite the increased complexity required to implement the memory and datapath power saving strategies. The peak performance at 5.0 V generates 60 Mpixels/sec of three RGB components with a 120 MHz clock frequency while dissipating 1.2 W [Gordon]. The subband decoder contains 415,000 transistors in a 9.5 mm x 8.7 mm area implemented in a 0.8 μm CMOS technology.

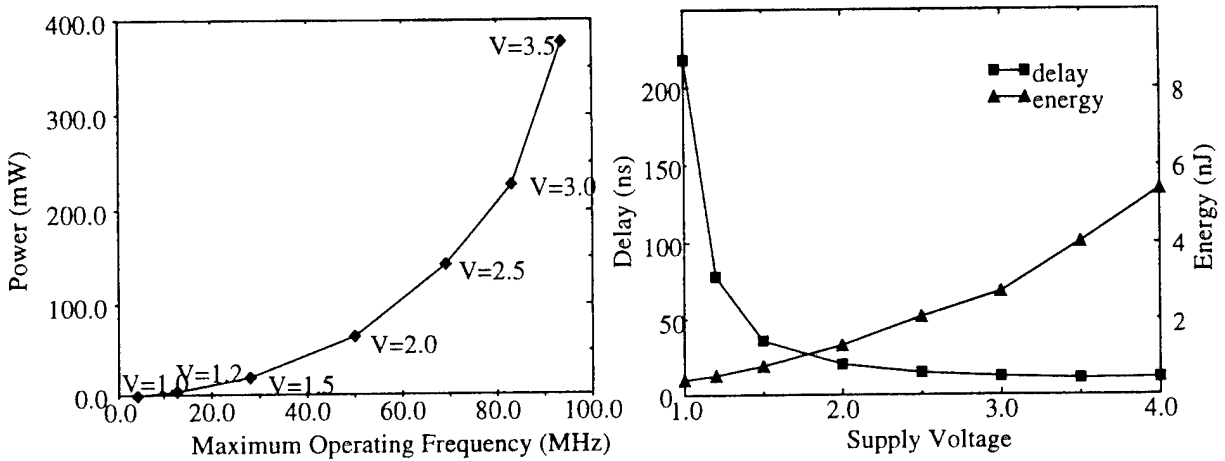


Figure 2. Performance measurements.

For higher-resolution images, multiple chips would be cascaded, each operating on a maximum of 256 pixels wide slice, producing a final image without boundary artifacts. The operating voltages are determined by the real-time computation requirements. This parallelism keeps the operating frequency and thus the supply voltage low, resulting in extremely low power dissipation (122 mW) even for HDTV applications.

The pyramid vector quantization (PVQ) decoder chip used together with the subband decoder chip for real-time video decompression was designed to operate at a 1.5 V supply and consume 6.6 mW at 6.4 MHz clock frequency, sufficient to decode 1.27 Mpixels/sec of color video with 160x240 pixels/frame at 30 frames/sec. The chip integrates 272K transistors and was implemented in 0.8mm triple-metal CMOS technology.

The complete video-on-demand system consists of a portable decoder with a color LCD display that decodes and displays compressed video sequences, a radio transmitter and receiver, and an encoding base station implemented in a DSP multiprocessor board. The wireless data transmission is provided by three pairs of direct-sequence spread spectrum radio transceivers manufactured by Proxim, delivering a raw data rate at 727 Kbits/sec. The decoding chip set on the portable decoder receives compressed video data and decompresses them to RGB color components, which are then converted to analog signals for the color display. The display is a 4" color thin film transistor active matrix display with a resolution of 160 pixels by 234 lines.

As shown in Fig. 3, the video-on-demand system accepts video data from two sources, a video server with a compressed video database and an NTSC camera. Video sequences stored in the video database are pre-compressed using our PVQ/subband encoding algorithm. As the compressed bit rate ranges from 0.5 Mbits/sec to 1 Mbits/sec, a simple bus interface between the video server (a SUN workstation) and the radio transmitter has been built to support this constant rate of data transfer.

The video-on-demand prototype system includes a real-time encoder to allow for live video sources such as a camcorder as an input device. The encoder consists of an NTSC decoder and multiple TMS400s on a DSP multiprocessor board, which implements the PVQ/subband encoding algorithm in real time. As our PVQ/subband algorithm is a symmetric compression algorithm, implying that the encoding and decoding procedures are almost identical, a low-power encoder would be feasible if a portable system of two-way video communication is desired.

From designing this prototype video-on-demand system, we learned that power reduction can be best attained through algorithm and architecture decisions, guided by the knowledge of underlying hardware and circuit properties. This hardware-driven algorithm design strategy is key to delivering high-quality video at an extremely low power level.

System Block Diagram

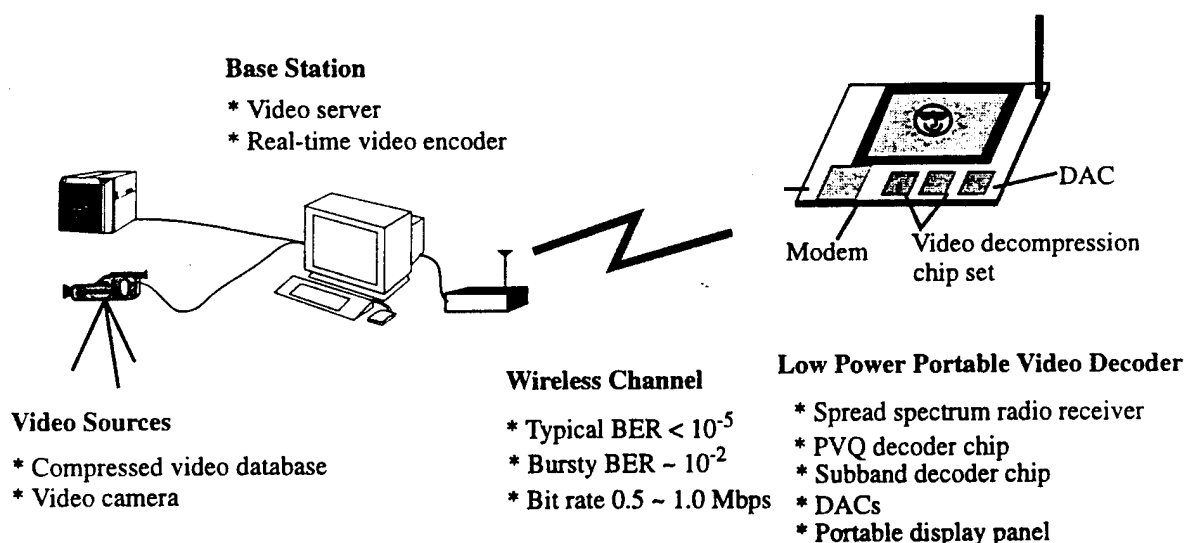


Figure 3. The portable video-on-demand system.

References

- [Chand] A. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low power CMOS digital design," *IEEE Journal of Solid-State Circuits*, pp. 685-691, April 1992.
- [Fisher] T. Fisher, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, vol. 32, pp. 568-583, July 1986.
- [Gordon] B. Gordon and T. Meng, "A 1.2mW video-rate 2D color subband decoder," *1995 IEEE Inter. Solid-State Circuits Conference Digest of Technical Papers*, pp. 290-291, February 1995.
- [Lee] W. Lee, *Mobile Cellular Telecommunication Systems*, McGraw-Hill, New York, 1989.
- [Purcell] S. Purcell, "C-Cube CL550 image processor," *Proc. HOT Chip Symposium*, August 1990.
- [Tseng] H. Tseng and T. Fisher, "Transform and hybrid transform/DPCM coding of images using pyramid vector quantization," *IEEE Trans. on Comm.*, vol. 35, pp. 79-86, January 1987.

JSEP Supported Presentation

1. T. H. Meng, "Portable Video-on-Demand in Wireless Communication," invited Forum presentation at Rockwell Science Center, March 1995.
2. B. M. Gordon, "A 1.2mW Video-Rate 2D Color Subband Decoder," presentation at 1995 *IEEE International Solid-State Circuits Conference*, February 1995.
3. B. M. Gordon, "A Low-Power Multiplierless YUV to RGB Converter Based on Human Vision Perception," presentation at 1994 *IEEE Workshop on VLSI Signal Processing*, October 1994.
4. T. H. Meng, "Wireless Video," invited presentation at the *Symposium on Video Communication*, May 1994.
5. T. H. Meng, "Low Power Video-on-Demand System," invited lecture to Watson Research Lab, IBM, May 1994.
6. T. H. Meng, "Portable Video-on-Demand in Wireless Communication," invited plenary talk at the *International Workshop on Low-Power Design*, April 1994.
7. T. H. Meng "Video Compression Algorithms for Portable Video-on-Demand," invited presentation at the *IEEE ISSCC Workshop on Multimedia*, February 1994.

JSEP Supported Publications

1. T. H. Meng, B. M. Gordon, E. K. Tsern, and A. C. Hung, "Portable Video-on-Demand in Wireless Communication," invited paper, to appear in the *Proceedings of IEEE*, April 1995.
2. B. M. Gordon, Teresa H. Y. Meng, "A 1.2mW Video-Rate 2D Color Subband Decoder," Digest of Technical Papers, 1995 *IEEE International Solid-State Circuits Conference*, pp. 290-291, February 1995.
3. B. M. Gordon, N. Chaddha and T. H. Meng, "A Low-Power Multiplierless YUV to RGB Converter Based on Human Vision Perception," *Proc. of 1994 IEEE Workshop on VLSI Signal Processing*, pp. 408-417, October 1994.
4. N. Chaddha, W. Tan, and T. H. Meng, "Fast Vector Quantization for Color Palette Design Based on Human Vision Perception," to appear in *IEEE Trans. on Image Processing*, 1995.
5. N. Chaddha, P. A. Chou, and T. H. Meng, "Scalable Compression based on Tree Structured Vector Quantization of Perceptually Weighted Generic Block, Lapped, and Wavelet Transforms," submitted to *International Conference on Image Compression*, October 1995.

Unit: 8

Title: ADAPTIVE DFE FOR GMSK IN INDOOR RADIO CHANNELS

Principal Investigator: J. M. Cioffi

Graduate Student: J. Tellado

Abstract

We simulate the performance of an equalized Gaussian Minimum Shift Keying (GMSK) signal in an indoor radio environment with fading and noise and Inter Symbol Interference (ISI). We show that data rates of 20 Mbps at Bit Error Rates (BER) $< 10^{-4}$ are possible with rms delay spreads up to 25 ns using a simple limiter-discriminator-integrator receiver and a (6,4) Decision Feedback Equalizer (DFE). In environments with larger rms delay spreads, coherent detection is required for the same performance. We introduce a DFE structure which compensates for both modulator and channel ISI, and yet requires no power-intensive multiplication operations in the feedback section. An (8,8) DFE with 2-level switched (selection) diversity is shown to allow 20 Mbps data transfer at $\text{BER} < 10^{-4}$ for rms delay spreads under 150 ns, with cochannel interference. Adding a (26,31) BCH code allows error-free reception of over 90% of packets with rms delay spreads under 150 ns, and up to 70% of packets with rms delays of 150 ns.

1. Introduction

Wireless Local Area Networks (LAN) support computing mobility on a local scale. They allow users to exchange information and retrieve files from the desktop or library as they walk about campus. High data rate radio links (on the order of 20 Mbps) will be needed to support the expected multimedia applications. Interest in these data rates has been reinforced by CEPT's allocation in Europe of 150 MHz of clear spectrum at 5.2 GHz for high data rate wireless LANs.

Adaptive equalization will be needed to overcome the severe ISI affecting transmission at such data rates. Clearly, low power consumption is critical in a portable radio communicator.

In addition to the ISI due to multipath, spectrally efficient GMSK introduces considerable amount of controlled phase ISI. Schemes to reduce this latter ISI and take advantage of simple noncoherent receivers have been studied [Ohno88]. Results have been published on mitigating multipath for linear modulations (QAM or QPSK), using only phase information [Ariya92] or amplitude and phase (e.g. [Rappaport93]). In [Steele94], a receiver is implemented to jointly reduce both ISI components, but is too complex for the proposed data rates.

In this paper, we quantify the equalization needed to transmit 20 Mbps using GMSK in radio environments with various levels of ISI impairment. We simulate the performance of an equalized GMSK signal in an indoor radio environment with fading and noise.

2. Channel Model

We use a n-ray model, with a baseband channel impulse response being defined as

$$h(t,k) = \sum_{i=1}^n a_i(k) d(t-iT_s/8),$$

where n is the effective number of paths, k is the channel number and $a_i(k)$ is the gain of the i -th path for channel k . There are eight impulse response taps per symbol period, T_s . In our simulation, the number of paths, n , varies between 33 and 194, depending on the value of the delay spread. The individual $a_i(k)$ are zero mean, complex Gaussian random variables, with an exponentially decaying profile. The resultant signal envelope has a Rayleigh distribution. The delay spread is adjusted by changing the exponent. The channels have been designed to have zero gain on average, but individual channels deviate from this because of Rayleigh fading.

Indoor channels change at pedestrian speeds of 1-2 m/s. Since the channel changes slowly relative to the symbol rate it is possible to transmit 20 kbit packets without tracking the channel after initial estimation.

Measurements show that most impulse responses have a σ under 100 ns in typical office buildings and open-plan factory buildings [Rappaport93]. To evaluate the performance of the different schemes, sets of 50 channels were generated for each of four σ s, 25, 50, 100 and 150 ns.

When the normalized delay spread, s_N , becomes greater than 0.1-0.2, some ISI mitigation method, such as equalization, must be applied to achieve Bit Error Rates (BER) under 10^{-3} , as shown in [Chuang87].

3. GMSK Receivers

In this section, we describe several GMSK receiver structures. We argue that the optimum detector is too complex for state-of-the-art DSP speeds, and explore suboptimum detection schemes.

Detectors

The optimal detector for CPM is a Maximum Likelihood Sequence Estimation (MLSE), which can be efficiently implemented using the Viterbi algorithm whenever the modulation index, $h=2k/p$ (k, p integers). However, the complexity of this algorithm grows exponentially with the memory of the modulation and of the channel.

Several proposed suboptimal structures use an equalizer to shorten the ISI, then apply a Viterbi type algorithm on the resulting shortened sequence, but the structure is still too complex for 20 Mbps at current DSP speeds. Alternatively, an adaptive equalizer can be used alone to mitigate the ISI in combination with a symbol by symbol GMSK detector such as: differential detection, frequency detection, and coherent detection.

Differential and frequency detection are simple to implement, but their performance degrades severely as BT decreases to values of interest for spectrally efficient radio communications ($.2 < BT < .3$). These non-coherent receivers disregard the amplitude of the received signal and therefore perform poorly for $s_N > .5$.

An adaptive equalizer can be used alone to mitigate the modulator and channel ISI. The results for several equalizer structures, are presented in the following section.

4. Adaptive Equalization

We first explore the performance impact of using symbol-spaced versus fractionally-spaced DFEs. Since the DFE-LDI (Limiter-Discriminator-Integrator) has shown to deliver acceptable performance only for $s < 25$ ns, we quantify the performance of a coherent receiver followed by a DFE. We end this section with a discussion of equalizer coefficients setting.

Simulation Assumptions

In this work, GMSK was simulated with a $BT=0.3$, as a compromise between spectral efficiency and added ISI. The data rate is 20 Mbps, and each packet is limited to 10 kbits, over which the stationary channel assumption holds. The feedback section of the DFE realistically uses the receiver decision outputs and thus the system suffers from error propagation.

Symbol-Spaced vs. Fractionally-Spaced DFE

Keeping the total number of taps constant, we found that the FS scheme has no performance advantage over the TS scheme.

DFE Taps

Since the radio channel varies on each packet, the (N_f, N_b) DFE weights need to be adapted for each incoming packet. In a computer network, variable-length packets often arrive in bursts. This LAN traffic pattern dictates that packets be detected in real time which requires $N_f + N_b$ complex multiplications and additions within T_s . Therefore minimizing the number of coefficients while maintaining acceptable performance is critical.

Decision Feedback Equalization for Coherent Detector

In this section, we explore the performance of a coherent detector followed by a DFE. This structure is more complex than the LDI-DFE, but is still implementable with present-day technology.

For $s=50$ ns, a (8,8) DFE provides error free 10kbit packets approximately 98% of the time. Using a larger number of feedback taps, such as (8,16), does not increase the performance when s is on the order of 50 ns or 100 ns. A DFE(8,16) is slightly better in channels with $s=150$ ns. We conclude that since the vast majority of indoor radio channels $s < 150$ ns, an DFE(8,8) is a good compromise between performance and complexity.

Computing Equalizer Coefficients

For $BT > .25$, the transmitter phase ISI can be truncated to 3 symbols and at the optimum sampling instant the phase becomes

$$f((n+1)T) = f_{\text{MSK}}((n+1)T) + (a_{n+1} - a_n) q_{BT},$$

For MSK, $q_{BT}=0$ and a coherent receiver suffers no transmitter ISI. For $BT=.3$, $q_{BT} = 16^\circ$ and the degradation is only about 1dB.

The sequence $\exp(jf((n+1)T))$ will take 12 different values in general. If we use this sequence to train the the equalizer, we will mitigate the channel ISI. If, on the other hand, we use $\exp(jf_{\text{MSK}}((n+1)T))$ as the training sequence, then the sequence can only take values $\{1, j, -1, -j\}$. This training sequence assumes that the transmitted sequence was actually MSK. This scheme does not require multiplications in the feedback section of the DFE, since multiplying any complex number by any of $\{1, j, -1, -j\}$ only involves swapping real and imaginary parts and sign changes. Now the feedback section of the DFE can be implemented using adders only.

Simulations show that the BER computed using either training sequences are equivalent. Hence we used $\exp(jf_{\text{MSK}}((n+1)T))$ due to its implementation advantages.

This proposed structure will realize $(4*N_f)$ real multiplies and $(4*N_f + 2*N_b - 3)$ real additions. The system trained with $\exp(jf((n+1)T))$ requires $(4*(N_f + N_b))$ real multiplies and $(4*(N_f + N_b) - 4)$ real additions. This translates into an almost 50% reduction of complexity and power consumption for $N_f = N_b$ and will be greater if $N_f < N_b$.

In noise-free channels with ISI, the coherent detector with DFE(8,8) is able to achieve error free transmission for all 25, 50 and 100ns channels and 96% of the 150ns channels when the equalizer coefficients converge to the optimum solution. This proposed structure is clearly effective in combatting distortion in ISI limited channels even for very long s_N .

Training Sequence Length

The equalizer must perform correctly across a large range of values of s . Using too many equalizer taps degrades the performance under finite training length. Finite training sequences lead to imperfect convergence, introducing noise. Using too few equalizer taps degrades the performance by not compensating for significant channel reflection paths.

The length of the training sequence should be as small a fraction of the packet size as possible to minimize overhead. The performance for LMS DFE(8,8) for training sequences of length 450 was found to be acceptable. Increasing the training sequence to 1000 from 450 only improved the performance slightly. The slow convergence of LMS was due to high eigenvalue spreads present in channels with deep fades. In contrast, RLS achieved LMS(450) performance with a 45 bit training sequence, and approached the asymptotic value at 100 bits.

5. Diversity

Diversity techniques are known to reduce the impact of radio channel fades. There are three main ways of processing the signals received from each diversity branch: Maximal Ratio Combining (MRC), Equal Gain Diversity (EGD), and Switched (or Selection) Diversity (SD). Unlike MRC and EGD, SD does not require co-phasing and operates with a single equalizer. Maximizing equalized decision point SNR would be an optimum SD criteria, but would require equalizing each branch signal. The simplest and most utilized technique receives on the branch with largest input $S+N+I$. We simulated selection diversity. With 2-level diversity we get 99%, 98%, 90% and 60% error free packets for the 25, 50, 100 and 150 ns channels.

Codes

Codes are of limited value when errors are bursty. However, our simulations show that a single-error-correcting BCH(26,31) code with a 31 by 16 matrix interleaver was effective in increasing error-free packet throughput at the output of the DFE. E.g the error free 10Kb packet throughput increased from 82% to 94% for the 100 ns channel and from 50% to 67% for the 150 ns case.

References

- [Ohno88] F. Adachi and K. Ohno, "Performance Analysis of GMSK Frequency Detection with DFE in Digital Land Mobile Radio," *IEE Proc.*, **Vol. 135**, Pt. F, no. 3, pp. 199-207, June. 1988.
- [Ariya92] S. Ariyavisitakul, "Equalization of Hard-Limited Slowly-Fading Multipath Signal Using a Phase Equalizer with Time-Reversal Structure," *IEEE J. Select. Areas Commun.*, **Vol. 10**, no. 3, pp. 589-598, April 1992.
- [Rappaport93] T. S. Rappaport, W. Huang and M. J. Feuerstein, "Performance of Decision Feedback Equalizers in Simulated Urban and Indoor Radio Channels," *IEICE Trans. on Commun.*, **Vol. E76-B**, no. 2, pp. 78-89, Feb. 1993.
- [Steele94] J. C. S. Cheung and R. Steele, "Soft-Decision Feedback Equalizer for Continuous Phase Modulated Signals in Wideband Mobile Radio Channels," *IEEE Trans. on Commun.*, **Vol. 42**, no. 2/3/4 pp. 1628-1638, 1994.
- [Chuang87] J. C-I Chuang, "The Effects of Time Delay Spread on Portable Radio Communications Channels with Digital Modulation," *IEEE J. on Select. Areas in Commun.*, **Vol. SAC-5**, no. 5, pp. 879-889, June 1987.
- [Murota81] K. Murota and K. Hirade, "GMSK Modulation for Digital Mobile Radio Telephony," *IEEE Trans. on Commun.*, **Vol. COM-29**, pp. 1044-1050, July 1981.
- [Aulin85] J. B. Anderson, T. Aulin and C. E. Sundberg, *Digital Phase Modulation*, Plenum Press, 1985.

Unit: 9

TITLE: Smart Antennas for Mobile Radio Networks

PRINCIPAL INVESTIGATOR: T. KAILATH

GRADUATE STUDENTS: Y.C. Pati and B. Hassibi

Exploiting Spatial Structure and Signal Structure in Smart Antennas for Mobile Radio Networks

1 Introduction

In this research our goal has been to investigate the use of antenna array processing algorithms in military and civilian mobile radio networks. We have found that using smart antennas at the network base station can substantially improve the capacity (number of users per cell) and quality (outage probability) of such networks.

Limiting factors in the capacity and quality of current wireless communication systems are the availability of RF frequency spectrum (which requires more efficient spectral usage) and the mutual interference between co-channel users. By using antenna arrays one can design smart transmitters that use information on the mobile unit locations to emit directional radiation towards the intended mobile unit while minimizing the radiated energy in the direction of other co-channel mobile units. Likewise, on receive, smart receivers can be designed that exploit location (or spatial) information to receive signals with full gain from the desired location while having very low (near zero) gain in the direction of other co-channel mobile units. Such techniques can result in greatly reduced mutual interference and may consequently allow multiple co-channel users within a single cell, thus boosting the capacity severalfold.

While popular techniques of direction-of-arrival (DOA) estimation (such as MUSIC and ESPRIT) have been successful in many applications, they have limitations in mobile communications environments where the number of users, and their corresponding multipaths, is typically much greater than the number of sensors. Fortunately, in communication systems, there are other resources available, such as a cooperative mobile unit and a known temporal structure for the signals. While several authors have investigated using training signals and signal structure information [Compton], their approaches have not attempted to combine all aspects of the temporal and spatial structures effectively.

Finding effective methods of combining spatial structure and temporal (signal) structure is a major challenge for mobile communications. In what follows we describe three approaches to addressing this problem. The first method combines spatio-temporal information for the blind identification of possibly non-minimum phase channels for multiple users. In the second and third methods are blind methods in which the signals are assumed to have a *constant modulus* structure which is the case for analog FM signals and for a great many digital modulation schemes (DPSK, QPSK, etc.). The second method uses higher-order statistics to estimate the array response matrix of the antenna array from which the original signals may then be recovered. The third method is a blind adaptive method that exploits the known bandwidth of the information signals together with the constant modulus property to separate and demodulate FM signals.

2 Spatio-Temporal Blind Identification of FIR Channels for Multiple Users

Equalization of a communications channel requires implicit or explicit knowledge of its transfer function. A communication channel is usually identified by LMS or RLS-type adaptive algorithms

in which the reference signal is provided by transmitting known training sequences. The so-called blind channel identification techniques only use the channel output and some known statistical properties of the transmitted signal. As a result, these techniques have the potential to increase the transmission capability by eliminating training sequences.

It is well known that nonminimum phase channels driven by wide-sense stationary input sequences, cannot be identified from second-order statistics. Therefore blind identification techniques, to date, use either higher order statistics [Tugnait, Shalvi, Porat, Hatzinakos] or use cyclostationary input signals [Gardner, Tonga, Tongb] to identify a possibly nonminimum phase channel. Due to their slow rate of convergence, these techniques may be impractical for mobile communications environments.

In general, signals arrive at the receiver not only with different delays, but also from different spatial angles. In digital transmission systems, antenna arrays have recently attracted much attention in the framework of optimal spatial diversity combining [Tongb, Balaban]. In [Khalaj] it was shown that for the antenna array case, second-order statistics provide enough information for identifying the channel, and a method was proposed on this basis.

In the multiple user case, second-order statistics are not sufficient to uniquely identify the channels for each user. Additional information or structure has to be exploited in order to obtain unique estimates of all the channel coefficients. In [Hassibia, Aghajan] we have used the spatial structure of the incoming signals. The spatial structure that we shall assume is that for *each user that exists some direction in which that user's contribution is dominant*.

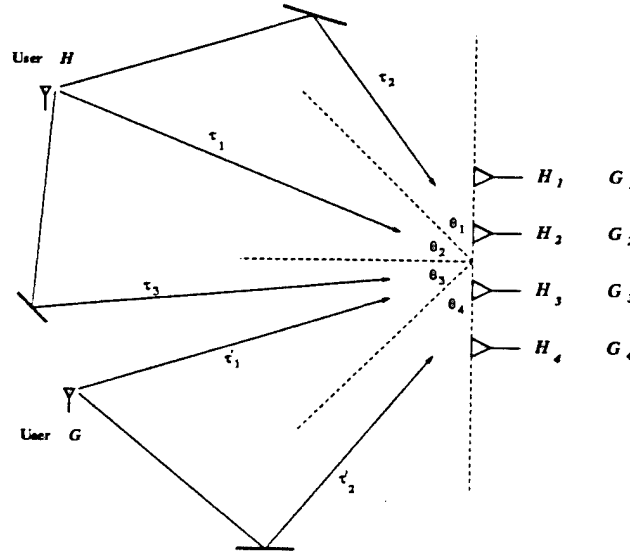


Figure 1: A typical multi-user channel model

To this end, consider Figure 1 where a typical propagation model is shown, in which four antenna elements receive signals from two users via multiple paths. The channel, for each user and for each antenna element, can hence be characterized in the time domain by an FIR filter with coefficients related to the ISI strengths. If we consider the channel induced by users H and G at the i th antenna element to be $h_i(t)$ and $g_i(t)$, respectively, then the signal received at the i th antenna element is seen to be

$$x_i(t) = \sum_{k=-\infty}^{\infty} s_k^1 h_i(t - kT) + \sum_{k=-\infty}^{\infty} s_k^2 g_i(t - kT) + n_i(t),$$

where s_k^1 and s_k^2 are the symbols transmitted by users H and G , and $n_i(t)$ is spatially and temporally

white noise with variance σ^2 . The output of an antenna array with M elements can be written in vector form as

$$\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T.$$

The power spectrum of $\mathbf{x}(t)$ is the z -transform of its autocorrelation matrix, viz.,

$$S_{\mathbf{x}}(z) = \mathcal{Z}(R_{\mathbf{x}}[m]) = \mathcal{Z}(E\mathbf{x}[k]\mathbf{x}^*[k-m]),$$

which under our assumptions can be written as

$$S_{\mathbf{x}}(z) = \begin{bmatrix} \mathbf{H}(z) & \mathbf{G}(z) \end{bmatrix} \begin{bmatrix} \mathbf{H}^*(z^{-1}) \\ \mathbf{G}^*(z^{-1}) \end{bmatrix} + \sigma^2 I \quad (1)$$

where

$$\mathbf{H}(z) = [H_1(z), H_2(z), \dots, H_M(z)]^T \text{ and } \mathbf{G}(z) = [G_1(z), G_2(z), \dots, G_M(z)]^T$$

are the multichannel transfer function vectors.

In the single user case $\mathbf{G}(z) \equiv 0$, and the channels in $\mathbf{H}(z)$ can be identified from (1) using the method of [Khalaj]. However, in the multiple user case $\mathbf{H}(z)$ and $\mathbf{G}(z)$ cannot be uniquely identified from (1), since if $\bar{\mathbf{H}}(z)$ and $\bar{\mathbf{G}}(z)$ satisfy (1), then so will

$$\begin{bmatrix} \mathbf{H}(z) & \mathbf{G}(z) \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}(z) & \bar{\mathbf{G}}(z) \end{bmatrix} \mathbf{U}$$

for any unitary matrix \mathbf{U} .

However, if we look at the problem in the spatial frequency domain we may transform the measurement vector $\mathbf{x}(t)$ to

$$\Theta(t) = \mathbf{W}\mathbf{x}(t)$$

where \mathbf{W} is the array response matrix at say M different spatial angles. This also results in the spatial power spectrum

$$S_{\Theta}(z) = \mathbf{W}S_{\mathbf{x}}(z)\mathbf{W} = \begin{bmatrix} \mathbf{H}_{\theta}(z) & \mathbf{G}_{\theta}(z) \end{bmatrix} \begin{bmatrix} \mathbf{H}_{\theta}^*(z^{-1}) \\ \mathbf{G}_{\theta}^*(z^{-1}) \end{bmatrix} + \sigma^2 I$$

where

$$\begin{bmatrix} \mathbf{H}_{\theta}(z) & \mathbf{G}_{\theta}(z) \end{bmatrix} = \mathbf{W} \begin{bmatrix} \mathbf{H}(z) & \mathbf{G}(z) \end{bmatrix}.$$

Now from our assumption, user H must be dominant in some direction, say θ_i , and user G must be dominant in some other direction, say θ_j . (Note that in Figure 1 user H is dominant in directions θ_1 and θ_2 , while user G is dominant in direction θ_4 .) Thus the channel response vectors $\mathbf{H}_{\theta}(z)$ and $\mathbf{G}_{\theta}(z)$ must have the following form

$$\begin{bmatrix} \mathbf{H}_{\theta}(z) & \mathbf{G}_{\theta}(z) \end{bmatrix} = \begin{bmatrix} \times & \times \\ \vdots & \vdots \\ \times & 0 \\ \vdots & \vdots \\ 0 & \times \\ \vdots & \vdots \\ \times & \times \end{bmatrix} \begin{matrix} \leftarrow \theta_i \\ \\ \leftarrow \theta_j \end{matrix} \quad (2)$$

It is readily observed from (2) that the only unitary matrix that will preserve the structure of the channel response vectors is

$$\mathbf{U} = \begin{bmatrix} e^{j\Omega_1} & 0 \\ 0 & e^{j\Omega_2} \end{bmatrix}.$$

Therefore, the channels are uniquely identifiable up to such a constant phase factor, and we may use the method of [Khalaj] to find the channel response vectors $\mathbf{H}_\theta(z)$ and $\mathbf{G}_\theta(z)$. The method is based upon forming a so-called *Sylvester matrix*, from estimates of the power spectrum, and computing its minimum singular vector.

The above method, which is described in more detail in [Hassibia, Aghajan], has applicability in the transmit antenna array (at the base station) where the multipath is not too severe and where the assumption of each user being dominant in some spatial direction is reasonable.

3 Closed-Form Solution to the Constant-Modulus Factorization Problem

In many mobile communications situations, such as in analog FM, or in the digital modulation schemes DPSK and QPSK, the transmitted signals have a constant modulus structure. As we shall presently see, this signal structure may be exploited to obtain estimates of the the array response matrix, without using training sequences, from which the original signals can be recovered.

To this end, consider an antenna array with M elements, and suppose $d < M$ independent constant modulus (CM) signals are received by this antenna array. Suppose, moreover, that the array manifold is not known. If we assume that the delay spreads between the multipaths are small (compared to the inverse of the signal bandwidth), then each CM signal will induce an array response vector $[a_{1i} \ a_{2i} \ \dots \ a_{Mi}]^T$ (which is unknown since we have assumed the array response vector unknown). If, for the time being, we assume that the noise is negligible, then the total response measured at the antenna array at time i is

$$\mathbf{x}_i = \mathbf{A} \mathbf{s}_i, \quad i = 1, \dots, N,$$

where

$$\mathbf{x}_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{Mi} \end{bmatrix}, \quad \mathbf{s}_i = \begin{bmatrix} s_{1i} \\ s_{2i} \\ \vdots \\ s_{di} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1d} \\ a_{21} & a_{22} & \dots & a_{2d} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \dots & a_{Nd} \end{bmatrix}, \quad d \leq M$$

and x_{ki} is the signal received at the k th antenna element at time i , and s_{li} is the the CM signal transmitted by user l at time i .

If we combine all the observed signals from time 0 to N in the matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \vdots & & \vdots \\ x_{M1} & x_{M2} & \dots & x_{MN} \end{bmatrix}, \quad N \gg M$$

and the unknown constant modulus signals into the matrix

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1N} \\ s_{21} & s_{22} & \dots & s_{2N} \\ \vdots & \vdots & & \vdots \\ s_{d1} & s_{d2} & \dots & s_{dN} \end{bmatrix}, \quad |s_{ij}| = 1$$

then we can write

$$\mathbf{X} = \mathbf{A} \mathbf{S}.$$

Problem 1 (CM Factorization Problem) Given such an X , find the factorization

$$X = AS,$$

where A and S have the aforementioned properties.

Note that we are interested in the factorization of Problem 1 since it will allow us to separate and identify each CM signal.

In the case when the noise is not negligible, X may not exactly admit a constant modulus factorization and all we can write is

$$X = AS + N,$$

for some noise matrix N . In this case one shall attempt to find a factorization of the form $\hat{X} = \hat{A}\hat{S}$, where \hat{X} is close to X in some sense.

For a long time, the constant modulus factorization problem was considered to be too nonlinear to admit a closed-form analytic solution, and only iterative gradient-descent schemes have been developed. These algorithms are based on the pioneering work of Godard [Godard] and Treichler and Agee [Treichler], and are intimately related to alternating projection methods. These algorithms go under the name CMA's (constant modulus algorithms) and have the drawback that convergence may be slow, or that one may converge to a local minimum.

In [Hassibib] we have shown that by making assumptions on the statistics of the phases of the signals, we can find a closed-form solution to the array response matrix A . The method is based on estimating the higher order statistics of the received signals and can be shown to yield asymptotically unbiased estimates. The method also allows one to find the A matrix in the presence of noise, and when an exact factorization does not exist.

More specifically, the method assumes knowledge of the higher-order moments

$$E[s_{ki}]^l, \quad l = 1, 2, \dots, d.$$

These are readily known for the following two cases of interest.

- (i) The $\{s_{ki}\}$ belong to a finite alphabet.
- (ii) $s_{ki} = e^{j\phi_{ki}}$, where the phases ϕ_{ki} are uniformly distributed over $[0, 2\pi]$. In this case

$$E[e^{j\phi_{ki}}]^l = 0, \quad \forall l.$$

Case (i) happens in digital modulation schemes and case (ii) is a reasonable assumption for FM signals.

Under these assumptions it is possible to find estimates of the polynomials

$$P^{(i)}(\alpha) = \alpha^d + p_1^{(i)}\alpha^{d-1} + \dots + p_d^{(i)} \quad i = 1, \dots, M$$

whose roots are

$$\{|a_{i1}|^2, |a_{i2}|^2, \dots, |a_{id}|^2\}$$

and the polynomials

$$P^{(ij)}(\alpha) = \alpha^d + p_1^{(ij)}\alpha^{d-1} + \dots + p_d^{(ij)} \quad i, j = 1, \dots, M$$

whose roots are

$$\{a_{i1}a_{j1}^*, a_{i2}a_{j2}^*, \dots, a_{id}a_{jd}^*\}.$$

The entries of the matrix A are then obtained by finding the roots of the polynomials $P^{(i)}(\alpha)$ and $P^{(ij)}(\alpha)$.

The above proposed method has the advantage that it avoids iteration and convergence problems, and that it is asymptotically unbiased. The feasibility of the method has been demonstrated using numerous computer simulations. It may also be used as an initial estimator for local gradient-based CM algorithms.

4 Blind Adaptive Demodulation of Co-channel FM Signals

To enhance the performance of blind methods for cochannel signal separation, it is important to exploit all available forms of *a priori* knowledge. The *constant modulus* methods mentioned above rely explicitly on the knowledge of only the constant modulus structure provided by the modulation format. However, in any given communications setting it is likely that there is additional information about signal structure that is ignored by CM methods. In the case of frequency modulated (FM) signals, as employed in the current analog cellular mobile systems, a key piece of information is the known bandwidth of the transmitted information signals.

The bandwidth of the information signal translates directly to the bandwidth of the phase $\phi(t)$, of the complex modulated carrier

$$s(t) = e^{j\phi(t)}.$$

This information is in fact currently exploited in the single signal case to design of phase-lock loop (PLL) demodulators that track the (slowly varying) phase of the received signal. In the single signal case, PLL's are a very effective means of exploiting the phase bandwidth of FM signals. In the case of multiple co-channel signals received at an antenna array we found that it is possible to combine the spatial model of the antenna array measurements i.e.

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N},$$

with the signal-structure (phase bandwidth) model employed in phase-lock loops to effectively separate and demodulate co-channel FM signals [Pati]. The method we propose is embodied in the architecture shown in Fig. 2, which we refer to as a *Multitarget Adaptive Phase-lock Loop* (MADPLL) demodulator. The constant modulus property of the FM signals is exploited by the bank of FM modulators shown. A key feature of the MADPLL demodulator is the simplicity of the architecture, which easily lends itself to real-time implementation.

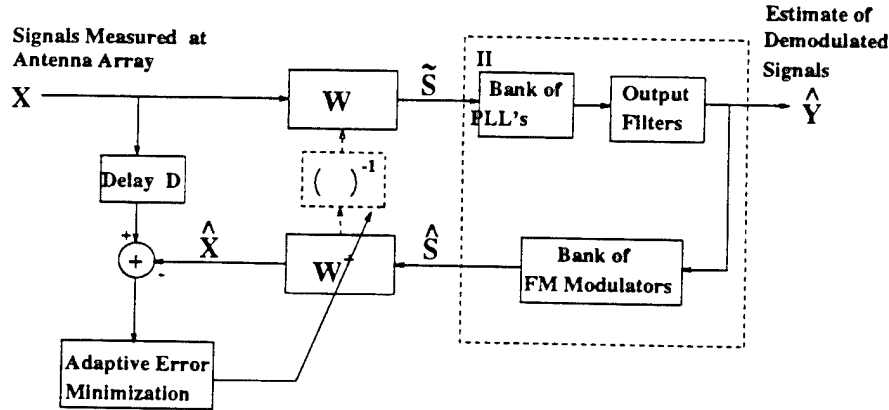


Figure 2: Block diagram of MADPLL structure for recovery of multiple cochannel FM signals.

The proposed MADPLL method may be described in the form of an adaptive algorithm as follows.

- (i) Estimate the modulated signals \mathbf{S} at time k using the current estimate \mathbf{W} for the inverse array response.

$$\tilde{\mathbf{S}}(k) = \mathbf{W}\mathbf{X}(k). \quad (3)$$

- (ii) Estimate the information signals $\hat{\mathbf{Y}}(k)$ by demodulating the $\tilde{\mathbf{S}}(k)$ using a bank of PLL's.

- (iii) Remodulate $\hat{\mathbf{Y}}(k)$ to obtain new estimates $\hat{\mathbf{S}}(k)$ for the frequency modulated signals.
- (iv) Estimate the signals received at the antenna array based on the current estimate $\hat{\mathbf{S}}(k)$, of the modulated signals and the current estimate \mathbf{W}^+ , of the array response, *i.e.*

$$\hat{\mathbf{X}}(k) = \mathbf{W}^+ \hat{\mathbf{S}}(k). \quad (4)$$

- (v) Update the estimated array response \mathbf{W} to decrease the error between the estimate $\hat{\mathbf{X}}$ and the measured \mathbf{X} , *e.g.* by minimizing

$$\mathcal{E}(k) = \sum_{l=0}^k \lambda^k \|\mathbf{X}(l) - \hat{\mathbf{X}}(l)\|^2, \quad 0 < \lambda \leq 1, \quad (5)$$

using the *recursive least-squares* (RLS) algorithm. (Any delay in the demodulation/remodulation step can be accounted for here with a corresponding delay in the error definition (see Fig. 2)).

- (vi) Update the inverse array response estimate \mathbf{W} by 'inverting' the array response $\mathbf{W}b f^+$. Here inversion implies psuedoinversion, $\mathbf{W} = (\mathbf{W}^+)^{\dagger}$.
- (vii) $k \leftarrow k + 1$, Go to Step (i).

The main difference between the MADPLL and the so-called constant modulus techniques is the use of "phase-smoothing" to incorporate the known bandwidth of the information signals. This provides additional leverage to the algorithm that improves both convergence speed and quality of the estimates.

To indicate the type of performance improvement over conventional CM methods that may be expected by exploiting the bandwidth of information signal in addition to the constant modulus property, consider the following simulation result. This simulation result is obtained for three sensors and three signals ($m = 3$, $d = 3$), with a sampling frequency $f_s = 180\text{kHz}$, carrier frequency $f_c = 60\text{ kHz}$, and signal bandwidth $f_b = 10\text{kHz}$. PLL FM demodulators were used to demodulate the signals, and the RLS algorithm was used to update the weight matrix.

In Fig. 3, the average output SNR's for the two cases are plotted as a function of average input CNR at the sensors. The output SNR's are computed using the output signal after a delay that accounts for the convergence time of the algorithm. It is observed that in general the proposed MADPLL method outperforms the use of CM properties alone. The actual amount of improvement is dependent on both the signal and the array response, and further analysis is required to establish bounds on performance.

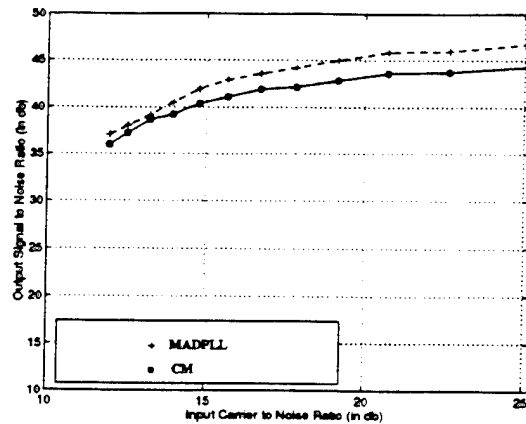


Figure 3: Average output SNR versus average input CNR for the constant gain channel. The averages are taken over 20 trials and over the three signals/sensors with random signals and array responses.

References

- [Aghajan] H. Aghajan, B. Hassibi, B. Khalaj, A. Paulraj and T. Kailath, Blind Identification of FIR Channels with Multiple Users via Spatio-Temporal Processing, in *Proceedings of GLOBECOM '94*, San Francisco, CA, Nov 1994.
- [Balaban] P. Balaban and J. Salz, Optimum Diversity Combining and Equalization in Digital Data Transmission with Applications to Cellular Mobile Radio - Part I: Theoretical Considerations, *IEEE Transactions on Communications*, **40(5)**:885-894, May 1992.
- [Compton] R.T. Compton, *Adaptive Antennas - Concepts and Performances*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [Gardner] W. Gardner, A new method of channel identification, *IEEE Transactions on Communications*, **39(6)**:813-817, June 1991.
- [Godard] D. N. Godard Self-recovering equalization and carrier tracking in two-dimensional data and communication systems, *IEEE Trans. on Communications*, vol. 28, pp. 1867-1875, Nov 1980.
- [Hassibia] B. Hassibi, H. Aghajan, B. Khalaj, A. Paulraj and T. Kailath, Spatio-Temporal Blind Identification of FIR Channels for Multiple Users, in *CISS '94*, Princeton, NJ.
- [Hassibib] B. Hassibi, A. Paulraj and T. Kailath, On a Closed Form Solution to the Constant-Modulus Factorization Problem, in *Proc. of 28th Asilomar Conf. on Circuits, Systems and Computers*, 1994.
- [Hatzinakos] D. Hatzinakos and C. Nikias, Estimation of multipath channel response in frequency selective channels, *IEEE J. Select. Areas Commun.*, **7(1)**:12-19, January 1989.
- [Khalaj] B. Khalaj, B. Hassibi, A. Paulraj, and T. Kailath, "Blind identification of FIR channels via antenna arrays", In *Proc. of 27 Asilomar Conf. on Circuits, Systems and Computers*, 1993.
- [Pati] Y. C. Pati, G. Raleigh, and A. Paulraj. Estimation of cochannel FM signals with multidimensional adaptive phase-locked loops and antenna arrays. In *Proceedings IEEE International Conference on Acoustics Speech and Signal Processing*, 1995. To appear.

- [Porat] B. Porat and B. Friedlander, Blind equalization of digital communication channels using high-order moments, *IEEE Trans. Signal Processing*, **39**:522-526, 1991.
- [Shalvi] O. Shalvi and E. Weinstein, New criteria for blind deconvolution of nonminimum phase systems, *IEEE Transactions on Information Theory*, **36**(2):312-320, March 1990.
- [Tonga] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels", In *Proc. of 25th Asilomar Conference on Circuits, Systems and Comp.*, 1991.
- [Tongb] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization of multipath channels: A Frequency domain approach", *IEEE Trans. on Information Theory*, **41**(1):329-334, January 1995.
- [Treichler] J. R. Treichler and B. G. Agee A new approach to multipath correction of constant modulus signals, *IEEE Trans. on ASSP*, vol. 31, pp. 456-471, Apr. 1983.
- [Tugnait] J.K. Tugnait, Identification of linear stochastic systems via second and fourth-order cumulant matching, *IEEE Transactions on Information Theory*, **33**:393-407, May 1987.

Publications

- [Siua] K-Y. Siu, V.P. Roychowdhury and T. Kailath, Rational approximation techniques for analysis of neural networks, *IEEE Trans. on Information Theory*, **40**(2):455-466, March 1994.
- [Siub] K-Y. Siu, A. Dembo and T. Kailath, On the perceptron learning algorithm on data with high precision, *J. of Comp. and System Sci.*, **48**(2):347-356, April 1994.
- [Siuc] K-Y. Siu, V. Roychowdhury and T. Kailath, Toward massively parallel design of multipliers, *J. of Parallel and Distributed Computing*, **24**:86-93, 1995
- [Xua] G. Xu, Fast subspace decomposition, *IEEE Trans. on ASSP*, **42**: 539-551, March 1994.
- [Xub] G. Xu, Y.M. Cho, and T. Kailath, Application of fast subspace decomposition to signal processing and communication problems, *IEEE Trans. on ASSP*, **42**(6):1453-1461, June 1994.
- [Xuc] G. Xu and T. Kailath, Fast estimation of principal eigenspace using Lanczos algorithm, *SIAM J. Matrix Anal. Appl.*, **15**(3):974-994, July 1994.
- [Xud] G. Xu, H. Zha, G. Golub and T. Kailath, Fast algorithms for updating signal subspaces, *IEEE Trans. on Circuits and Systems, II*, **41**(8):537-549, Aug. 1994.

TITLE: Efficient Data Compression

PRINCIPAL INVESTIGATOR: T. Cover

GRADUATE STUDENTS DIRECTLY SUPPORTED BY JSEP: L. An, T. Chung, E. Erkip, P. Fahn, D. Kimber, I. Kontoyiannis, S. Langford, A. Lapidoth, E. Ordentlich

AFFILIATED STUDENTS: V. Castelli, T. Jing, G. Iyengar

1 Scientific Objectives

We apply techniques of information theory to problems of information compression, image compression, distributed data compression and storage, and network information flow. We also investigate techniques for lossy and noiseless data compression. Our focus is on arithmetic coding for noiseless compression and on a new approach to lossy compression motivated by Kolmogorow complexity. We are particularly interested in the sequential refinement of images, where one sends information which can be refined in stages so that it is optimal at each step. During the past year this work resulted in 7 supported papers and 3 Ph.D. theses.

2 Summary of Research

Several research objects are bearing fruit. They involve data compression for images, an experiment to find the rate distortion function for images, the duality of image compression and investment, the role of pattern recognition in data compression, and methods for modeling voice and voice classification.

In [Castelli and Cover], we report new results which quantify the relative value of labeled and unlabeled samples for the classification of an unknown sample. We show that labeled samples are exponentially more valuable than unlabeled samples in pattern classification. This is shown in a series of two papers. The first paper considers the problem in the parametric case and the second when the underlying distributions are unknown but are otherwise identifiable through an infinite number of samples. Ordentlich[Ordentlich] has extended the work of Pombra [Pombra and Cover] on finding the capacity region of a multiple access Gaussian channel with nonwhite additive Gaussian noise. It is shown that capacity is improved by at most a factor of two.

3 Detailed Research Descriptions

3.1 Image Compression

Our experiment on finding the rate distortion function for images is being run by Tai Jing and Li An. A good example of our approach is the following. Fifteen years ago the best chess programs played a good game of chess, but it was known that grand masters could easily beat them. Thus, grand masters provided an existence proof that computer chess algorithms could be improved.

We take the same point of view now for image compression. There are existing algorithms, JPEG for example, which compress images. However, it is clear that computers can't 'see', while humans can. Thus, we expect that humans can compress images better than computers, and we want to know how much better. So we have designed an experiment for noisy image compression

which tries to show that, say at a compression of 60:1, that the images humans are able to describe at this rate are much superior to those of existing algorithms.

We are achieving initial success. The comparison of two images at 60:1 data compression, one generated by a human data compression, the other by JPEG, is striking. On the other hand, we're getting some indication that sub-band coding techniques are doing quite well with respect to human performance. We shall see what the final conclusions are.

We are interested in finding an accurate estimate of the minimal rate at which an image can be coded without incurring significant perceptible distortion. We do this by first having one experimental subject simplify a given image without significantly distorting it, and then have another subject predict the simplified image, pixel by pixel, as accurately as possible.

The accuracy of the second subject's predictions can be quantified to yield an estimate of the entropy of the simplified image. This two-stage process — simplification followed by noiseless data compression — can be mechanized and can be proven to be optimal in a rate distortion sense. Thus not only will our estimate of the minimum achievable rate with little or no perceptible distortion be useful as a benchmark to researchers in the field, but our experimental framework itself may pioneer a new algorithm for data compression. Our bound on the achievable compression should show that substantial improvement over current schemes is possible.

As far as details of the experiment are concerned, we plan to use human beings to 'cartoon' or simplify the existing image into one which is very predictable by another human being. This simplified image must also be psychologically nearby in that it is hard for a human observer to distinguish it from the original. This distortion measure is not quantitative, but it is clear that when the two images are almost indistinguishable, the lossy compression scheme is working well.

We will ask a person who has not previously seen the original or simplified image to predict at each pixel what shades are likely to appear in the next pixel given the pixels that have been seen so far. This person will then bet in these proportions on what shade will be next. It can be proved that if the subject achieves a wealth of 2^k starting from a one unit bet on the image, then k bits can be saved in the descriptive complexity of that image. Thus, an n -pixel image will require $n - k$ bits to describe. This is essentially the method used in [Cover and King] to find the entropy of English text. (It was found that a noiseless data compression of English text of a factor of 4:1 could be achieved, and it was argued that no further reduction was possible.)

3.2 Feedback in Communication

In [Pombra and Cover], it was shown that the feedback capacity of a non-white additive Gaussian noise channel can be evaluated via the maximization of the determinant of a certain covariance matrix under a trace constraint. We have developed an algorithm for maximizing this determinant and conjecture that the values attained are indeed global maxima.

Ordentlich has considered the optimum strategy for communicating with feedback in the presence of Gaussian noise. Preliminary results indicate that the majority of one's available power should be used to transmit linear combinations of past noise samples (which are available as a result of feedback) in an effort to reshape the effective noise spectrum. Specifically one should decrease the effective noise power where it is already low at the expense of increasing it where it is high.

It was recently shown by [Pombra and Cover] that the maximum achievable throughput (sum of rates of all users) of a Gaussian multiple access channel with feedback is at most twice that achievable without feedback. We prove [Ordentlich] a somewhat stronger result which establishes the factor of two bound not only for the total throughput but for the entire capacity region as well. Specifically, we show that the capacity region of a Gaussian multiple access channel with feedback is contained within twice the capacity region without feedback.

3.3 The Relative Value of Labeled and Unlabeled Samples

In [Castelli and Cover], Vittorio Castelli presented new results on the relative value of labeled and unlabeled samples in reducing the probability of error of the classification of a sample into one of two classes based on the previous observation of labeled and unlabeled data. Castelli and Cover showed that when the training set contains an infinite number of unlabeled samples, the first labeled sample reduces the probability of error to within a factor of two of the Bayes risk. The Bayes risk is the best that a classifier could do with complete a-priori knowledge of the densities associated with the two classes and the probability of each class. Further, they showed that subsequent labeled samples yield exponential convergence of the probability of error to the Bayes risk. Finally they show that labeled samples are exponentially more valuable than unlabeled samples with the relevant exponent being the Bhattacharya distance.

3.4 Robustness of Communication

Lapidoth, in a series of papers [Lapidoth1][Lapidoth2][Lapidoth3], has considered the robustness of signaling in the presence of noise in an unknown environment. It is well known that Gaussian signals and matched filter decoding is optimal for signaling with a power constraint over an additive white noise Gaussian channel. This is the basis for much of the signaling which is done, say, in deep space communication or in mobile communication. Lapidoth is able to show that even if you fix the receiver to be a matched filter receiver and continue to use the same signals, the information will get through the channel no matter what the noise is, just so long as the total noise power is not increased. Specifically, if the distribution of the noise is changed from Gaussian and independent to non-Gaussian and arbitrarily time dependent, as long as the noise power is not increased, the channel will still work and the probability of error will be exponentially small. This is a powerful real result showing the robustness of existing communication schemes to changes in the underlying assumptions on the model. We hope to generalize this result to non-Gaussian channels.

References

- [Castelli and Cover] V. Castelli and T. Cover. On the Exponential Value of Labeled Samples. *Pattern Recognition Letters*, 16:105-111, January 1995.
- [Cover and King] T. Cover and R. King. A Convergent Gambling Estimate of the Entropy of English. *IEEE Trans. on Information Theory*, IT-24(4):413-421, July 1978.
- [Lapidoth1] S. Shamai and Amos Lapidoth. Bounds on the Capacity of a Spectrally Constrained Poisson Channel. *IEEE Transactions on Information Theory*, 39(1):19-29, January 1993.
- [Lapidoth2] Amos Lapidoth. On the Reliability Function of the Ideal Poisson Channel with Noiseless Feedback. *IEEE Transactions on Information Theory*, 39(2):491-503, March 1993.
- [Lapidoth3] Amos Lapidoth. The Performance of Convolutional Codes on the Block Erasure Channel Using Various Finite Interleaving Techniques. *IEEE Transactions on Information Theory*, 40(5):1459-1473, September 1994.
- [Ordentlich] E. Ordentlich. A Class of Optimal Coding Schemes for Moving Average Additive Gaussian Noise Channels with Feedback. *Proceedings of the IEEE International Symposium on Information Theory*, p.467, June 1994.
- [Pombra and Cover] S. Pombra and T. Cover. Non-White Gaussian Multiple Access Channels with Feedback, *IEEE Transactions on Information Theory*, 40(3):885-892, May 1994.

4 Publications Supported by JSEP

4.1 Ph.D. Theses Supported by JSEP

The three Ph.D. theses supported this year are:

- D. Kimber, "Geometric Methods for Nonparametric Modeling of Dynamical Systems," December 1994;
- T. Chung, "Minimax Learning in Iterated Games via Distributional Majorization" June 1994;
- V. Castelli, "The Relative Value of Labeled and Unlabeled Samples in Pattern Recognition," December 1994.

4.2 Published Papers Supported by JSEP

1. Amos Lapidoth. On the Reliability Function of the Ideal Poisson Channel with Noiseless Feedback. *IEEE Transactions on Information Theory*, 39(2):491-503, March 1993.
2. L. Ekroot and T. Cover. The Entropy of Markov Trajectories, *IEEE Transactions on Information Theory*, 39(4):1418-1421, July 1993.
3. S. Pombra and T. Cover. Non-White Gaussian Multiple Access Channels with Feedback, *IEEE Transactions on Information Theory*, 40(3):885-892, May 1994.
4. Z. Zhang and T. Cover. On the Maximum Entropy of the Sum of Two Dependent Random Variables. *IEEE Transactions on Information Theory*, 40(4):1244-1246, July 1994.
5. V. Castelli and T. Cover. On the Exponential Value of Labeled Samples. *Pattern Recognition Letters*, 16:105-111, January 1995.

4.3 Papers Submitted for Publication

1. A. Lapidoth. Mismatched Decoding and the Multiple-Access Channel. *Stanford University Statistics Department Technical Report No.87*, February 1995.
2. T. Cover and E. Ordentlich. Universal Portfolios with Side Information. Submitted to *IEEE Transactions on Information Theory*.